P1 1175590

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

May 27, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/476,408*
FILING DATE: *June 05, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/09215*

REC'D 0 1 JUN 2004

WIPO   PCT

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

M. SIAS
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

06-09-03

| Atty. Docket No.: PF-1688 P | Type a plus sign (+) inside this box ➡ | + |

PROVISIONAL APPLICATION FOR PATENT COVER SHEET
(Large Entity)

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2).

TITLE OF THE INVENTION (500 characters max)	
KINASES AND PHOSPHATASES	
INVENTOR(S)/APPLICANT(S)	
FIRST NAME/MIDDLE INITIAL/LAST NAME	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Henry Yue	Sunnyvale, California
Danniel B. Nguyen	San Jose, California
Michael B. Thornton	Oakland, California
Rajagaopal Gururajan	San Jose, California
Ameena R. Gandhi	San Francisco, California
Yan Lu	Mountain View, California
Monique G. Yao	Carmel, Indiana
Mariah R. Baughn	San Leandro, California
Narinder K. Chawla	Union City, California
Joana X. Li	Millbrae, California
Wen Luo	South San Francisco, California
Ernestine A. Lee	Castro Valley, California
Ian J. Forsythe	Edmonton, Alberta Canada
Shanya D. Becha	San Francisco, California
Craig H. Ison	San Jose, California

27904

| Atty. Docket No.: *PF-1688 P* | *Type a plus sign (+) inside this box* ➡ | ✚ |

CORRESPONDENCE ADDRESS
(Customer No.: 27904)

| INCYTE CORPORATION
Legal Department
3160 Porter Drive
Palo Alto, California 94304 | Phone: (650) 855-0555
Fax: (650) 849-8886 or
(650) 845-4166 |

ENCLOSURES

Enclosed are:

1. Return Receipt Postcard;

2. Provisional Application Cover Sheet w/certificate of Express Mailing (2 pp., in duplicate);

3. __97__ Pages of Specification   (1-97)   ;

4. __9__   Pages of Claims   (98 -106)   ;

5. __1__   Page of Abstract   (107)   ;

6. __29__ Pages of Tables   (Tables 1 -8)   ;

7. __20__ Pages of Sequence Listing   (1-20)   ;

METHOD OF PAYMENT

Applicants hereby authorize the Commissioner to treat any concurrent or future reply in this application that requires a petition for an extension of time under 37 CFR 1.136(a) to be timely, as incorporating a petition for extension of time for the appropriate length of time; and to charge all required fees, including fees under 37 CFR 1.16, 1.17, and all required extension of time fees, or to credit any overpayment, to Deposit Account 09-0108. This sheet is enclosed in duplicate.

PROVISIONAL FILING FEE AMOUNT: $ 160.00

The invention was not made by an agency of the United States Government or under a contract with an agency of the United States Government.

Date: ___Jone 5, 2002___

Respectfully submitted,

INCYTE CORPORATION

Joel Harris
Reg. No. 44,743
Direct Dial Telephone: (650) 845-4866

KINASES AND PHOSPHATASES

TECHNICAL FIELD

The invention relates to novel nucleic acids, kinases and phosphatases encoded by these

5   nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and

prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders

affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. The

invention also relates to the assessment of the effects of exogenous compounds on the expression of

nucleic acids and kinases and phosphatases.

10

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the

intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in

a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate

15   groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues

serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups.

Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors

can activate kinases, which can occur as cell surface receptors or as the activator of the final effector

protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur,

20   as well as kinases sensitive to second messenger molecules. This system allows for the amplification

of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of

many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining

the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes

such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and

25   cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target

molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to

a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most

30   kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of

molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are

phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the

acceptor molecule, causing internal conformational changes and potentially influencing

intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating

35   protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

1

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

5   Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked

10   to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

        There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine

15   residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to

20   the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence

25   motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

        In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino-acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press,

30   San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

35   **Protein Tyrosine Kinases**

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second

5 messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that

10 function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied

15 by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs

20 are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell

25 proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases,

30 and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and

35 ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates

its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase

domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second

messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol,

5   3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The

PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP

produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of

hormone action in all animal cells that have been studied. Hormone-induced cellular responses

include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown,

10   bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in

all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA

expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders,

diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's

Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

15   The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein

kinases. This continuously expanding group of kinases have been implicated in the regulation of

numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and

repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of

eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol.

20   Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a

highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that

ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem.

273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification

25   of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms,

α, β, γ, δ, and ε. Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic

protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was

determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor

CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype

30   caused by deletion of the yeast CKI locus, HRR250 (Fish et al., supra).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of

CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau

locus is encoded by casein kinase I-epsilon, which is also a homolog of the Drosophila circadian gene

double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the

35   mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore

5 the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127).

10 HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomian transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of

15 which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus

20 (Nishiyama, Y. et al. (1999) FEBS Lett. 459:159-165).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium

25 receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-

30 3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and seratonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be

35 enriched in mammalian forebrain. This protein is associated with vesicles in both axons and

dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

5   The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase

10   cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang,X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun

15   N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

20 ## Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the

25   mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation

30   of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome

35   segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint

pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in

5 yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of

10 Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation

15 of the cell cycle and cell proliferation in human megakarocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

20 ### 5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses

25 such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

30 ### Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a

35 C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This

C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share

5   homology with the death-associated protein kinases (DAP kinases), known to function in interferon-γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the

10   N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a

15   member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal

20   kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

25   **Mitochondrial Protein Kinases**

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-

30   dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory

35   enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to

a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes

5 in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases

10 involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the

15 plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP$_2$). PIP$_2$ is then cleaved into inositol triphosphate (IP$_3$) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle

20 contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid

25 moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP$_2$) to PI (3,4,5) P$_3$ (PIP$_3$). PIP$_3$ then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins,

30 Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of

35 D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has

PF-1688 P

emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama,
T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein
coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP
regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by

5   sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of
sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including
platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C,
increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive
inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al.,

10   *supra*).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and
guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism
and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These

15   two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary
source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of
various steps in the synthesis of these two molecules has been the basis of many antiproliferative
drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of

20   ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated
with mitochondria and myofibrils, the subcellular structures that are involved in energy production
and utilization, respectively. Recent studies have demonstrated a major function for AdK in
transferring high energy phosphoryls from metabolic processes generating ATP to cellular
components consuming ATP (Zeleznikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus

25   AdK may have a pivotal role in maintaining energy production in cells, particularly those having a
high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its
activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of
various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be
treatable by increasing AdK activity.

30   GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA
synthesis, also fulfills an essential function in signal transduction pathways of cells through the
regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates
the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production
of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and

35   GTP levels also control the activity of certain oncogenic proteins such as p21[ms] known to be involved

in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21^{ras} and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

5 · GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for

10 reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of

15 deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of · DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Amer E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

PHOSPHATASES

20 Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated · · hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle

25 · and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous-structures-in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+}, for activity. PSPs play important roles in glycogen metabolism, muscle

30 contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE

35 PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular

localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A

5   number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) Curr. Opin.

10 Neurobiol. 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) Otolaryngol. Head Neck Surg.121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit

15 additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) Trends Biosci. 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE

20 PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through

25 the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases in vitro, and other evidence suggests that the same is true in vivo for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases,

30 cyclin-dependent kinases, and the IκB kinases (reviewed in Millward et al., supra). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the

35 phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been

proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

PP2B, or calcineurin, is a Ca^{2+}-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the

5   calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates

10   functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide

15   repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent

20   cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

25   In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site

30   consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently

35   affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain

fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte

5   activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2

10 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990)

15 Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The

20 MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine

25 phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Apyrases are enzymes that efficiently hydrolyze ATP and ADP and may function either intra- or extracellularly. One type of apyrase, ATP-diphosphohydrolase, catalyzes the hydrolysis of

30 phosphoanhydride bonds of nucleoside tri- and di-phosphates in the presence of divalent cations (Nourizad, N. et al., (2003) Protein Purif. 27:229-237).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C.

35 The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular

signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine

residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated

in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and

apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996)

5   J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid

phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and

maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri,

D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are

associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst.

10   91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase,

hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters

and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each

centered around a histidine residue which is involved in catalytic activity. Members of the HAP

15   family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both

sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at

positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-

coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein,

20   EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-

Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the

EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally,

synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane,

and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-

25   320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to

support coat formation of endocytic vesicles more effectively than was seen in wild-type mice,

suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These

findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic

vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

30   Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of

molecules spatially distributed over, and stably associated with, the surface of a solid support.

Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in

a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping,

35   bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants.

5   When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Lung Cancer

10   The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of cancer, such as lung cancer. Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent

15   DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of

20   diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous

25   cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group

30   accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions

35   associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of

this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as

5   RB, p53 and CDKN2.

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda et al. (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down

10   regulated whereas nm23H1 was upregulated. Petersen et al. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes, thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang et al. (2000; Oncogene

15   19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpresssed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakofillin 1 and cytokeratin 13.

There is a need in the art for new compositions, including nucleic acids and proteins, for the

20   diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

SUMMARY OF THE INVENTION

25   Various embodiments of the invention provide purified polypeptides, kinases and phosphatases, referred to collectively as 'KPP' and individually as 'KPP-1,' 'KPP-2,' 'KPP-3,' 'KPP-4,' and 'KPP-5' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for

30   facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a

35   polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-

5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide

5   having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-5.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring

10   amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In another embodiment, the polynucleotide encodes a

15   polypeptide selected from the group consisting of SEQ ID NO:1-5. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:6-10.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting

20   of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

25   Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least

30   90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide,

35   wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence

18

operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

5   sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

10   selected from the group consisting of SEQ ID NO:1-5.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected

15   from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample,

20   said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide

25   complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b)

30   detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide

35   comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a

polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method

5   comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a

10   polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ

15   ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional KPP, comprising

20   administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an

25   amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample.

30   Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness

35   as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an

amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

5 consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment

10 provides a method of treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a

15 naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) combining the

20 polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

25 acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

30 selected from the group consisting of SEQ ID NO:1-5. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide

35 in the presence of the test compound is indicative of a compound that modulates the activity of the

polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, the method

5 comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said

10 method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at

15 least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting

20 of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-

25 iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank

35 homolog, and the PROTEOME database identification numbers and annotations of PROTEOME

database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the

5 polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of

10 the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

15

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is

20 for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one

25 or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now

30 described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

5   The term "agonist" refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

. An "allelic variant" is an alternative form of the gene encoding KPP. Allelic variants may

10   result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times

15   in a given sequence.

"Altered" nucleic acid sequences encoding KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe

20   of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding KPP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity,

25   charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids

30   with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

. The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally

35   occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino

acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other

5   nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway

10   in which KPP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')$_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide

15   used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

20   makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

25   The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include

30   deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH$_2$), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

35   Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a

cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

5   The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense"
10   (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-
15   deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

20   The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

25 .   "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The
30   composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's
35   solution, dry milk, salmon sperm DNA, etc.).

26

PF-1688 P

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer

5 program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of

10 the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide

35 backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

40 The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which

retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a
5   measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

10   "Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of KPP or a polynucleotide encoding KPP which can be
15   identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino
20   acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the
25   present embodiments.

A fragment of SEQ ID NO:6-10 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:6-10, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:6-10 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and
30   amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotides. The precise length of a fragment of SEQ ID NO:6-10 and the region of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-5 is encoded by a fragment of SEQ ID NO:6-10. A fragment of
35   SEQ ID NO:1-5 can comprise a region of unique amino acid sequence that specifically identifies SEQ

ID NO:1-5. For example, a fragment of SEQ ID NO:1-5 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-5. The precise length of a fragment of SEQ ID NO:1-5 and the region of SEQ ID NO:1-5 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods

5   described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or

10   more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two

15   sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the

20   LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

25   Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence

30   analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST

35   programs are commonly used with gap and other parameters set to default settings. For example, to

PF-1688 P

compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version

2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

> *Matrix: BLOSUM62*
>
> *Reward for match: 1*

5 *Penalty for mismatch: -2*

> *Open Gap: 5 and Extension Gap: 2 penalties*
>
> *Gap x drop-off: 50*
>
> *Expect: 10*
>
> *Word Size: 11*

10 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous

15 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes

20 in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some

25 alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions,

30 between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e

35 sequence alignment program (described and referenced above). For pairwise alignments of

30

PF-1688 P

polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise
5   comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

10   *Gap x drop-off: 50*

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence,
15   for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be
20   used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid
25   sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of
30   complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable
35   by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas

31

wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

5   Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and

10   conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS,

15   for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular

20   circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by

25   virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

30   The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect

35   cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the

5   art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

10   The term "modulate" refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or

15   synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

20   sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition.

25   PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will

30   vary by cell type depending on the enzymatic milieu of KPP.

"Probe" refers to nucleic acids encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are

35   short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide

33

by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

5   Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

10   Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

          Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 20   100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from 25   megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection 30   programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both 35   unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and

34

polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

5   A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term

10   recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a

15   vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription,

20.   translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

25   An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing KPP,

30   nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or

35   synthetic binding composition. The interaction is dependent upon the presence of a particular

structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

5   The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

   A "substitution" refers to the replacement of one or more amino acid residues or nucleotides 10 by different amino acid residues or nucleotides, respectively.

   "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

15   A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

   "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid 20 sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well 25 as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

   A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor 30 of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms 35 contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants

PF-1688 P

and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

5   A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least

10   93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide

15   may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide

20   polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a

25   certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one

30   of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or

35   prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders

37

affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is

5 denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

10 Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID

15 NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

20 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and

25 potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:3 is 99% identical, from residue M1 to residue K487, to human apyrase (GenBank ID

30 g4583675) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.7e-264, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also has homology to proteins that are localized to the Golgi, have lysosomal/autophagic function, and are apyrase proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:3 also contains a GDA1/CD39

35 (nucleoside phosphatase family) domain as determined by searching for statistically significant

matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:3 is a nucleoside phosphatase. SEQ ID NO:1-2 and SEQ ID NO:4-5 were analyzed and annotated in a

5   similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-5 are described in Table 7.

As shown in Table 4, full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number

10   (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 lists fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:6-10 or that distinguish between SEQ ID NO:6-10 and related polynucleotides. Column 3 shows identification numbers

15   corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide embodiments. Columns 4 and 5 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 3 relative to their respective full length sequences.

20   The identification numbers in Column 3 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 5026615F6 is the identification number of an Incyte cDNA sequence, and COLCDIT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 56087809H1). Alternatively, the identification numbers in column 3

25   may refer to GenBank cDNAs or ESTs (e.g., g22918955) which contributed to the assembly of the full length polynucleotides. In addition, the identification numbers in column 3 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (i.e., those sequences including the designation "ENST"). Alternatively, the identification numbers in column 3 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (i.e., those sequences

30   including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (i.e., those sequences including the designation "NP"). Alternatively, the identification numbers in column 3 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_XXXXXX_N$_1$_N$_2$_YYYYY_N$_3$_N$_4$ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the

35   algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and

PF-1688 P

$N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 3 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example,

FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with

5   *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a

10   RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

     Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis

15   methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

     In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA

25   identification numbers are not shown.

     Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to

30   construct the cDNA libraries shown in Table 5 are described in Table 6.

     Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show

the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence

5 at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected)

10 indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses KPP variants. Various embodiments of KPP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the KPP amino acid sequence, and can contain at least one functional or structural characteristic of KPP.

15 Various embodiments also encompass polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10, which encodes KPP. The polynucleotide sequences of SEQ ID NO:6-10, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar

20 backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding KPP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence

25 selected from the group consisting of SEQ ID NO:6-10 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6-10. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

30 In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding KPP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding KPP, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively

35 less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a

polynucleotide encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding KPP. Any one of the splice variants described above can encode a polypeptide which contains at least one

5   functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide

10   sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode KPP and its variants are generally capable of

15   hybridizing to polynucleotides encoding naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular

20   codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode KPP and KPP

25   derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding KPP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of

30   hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:6-10 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

35   Methods for DNA sequencing are well known in the art and may be used to practice any of

42

the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification

5   system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other

10   systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as

15   promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known

20   genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence

25   before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available

30   software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

35   sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T)

library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary

5 sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer

10 controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the

15 genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express KPP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by

20 random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such

25 as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene

30 variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected-to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random

35 point mutations may be recombined, screened, and then reshuffled until the desired properties are

44

optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

5   In another embodiment, polynucleotides encoding KPP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase

10   techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a

15   polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, supra, pp. 28-53).

20   In order to express a biologically active KPP, the polynucleotides encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding

25   KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational

30   control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994)

35   Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding KPP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al.,

5   *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression

10   vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-

15   1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue,

20   or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending

25   upon the use intended for polynucleotides encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding KPP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria-

30   containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or

35   T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable

5   integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of KPP. Transcription of polynucleotides encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used

10   alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

15   transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite

20   leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses KPP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

25   Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression

30   of KPP in cell lines is preferred. For example, polynucleotides encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to

35   confer resistance to a selective agent, and its presence allows growth and recovery of cells which

47

successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine

5 phosphoribosyltransferase genes, for use in tk^- and apr^- cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et

10 al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used.

15 These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995). Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the

20 sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

25 In general, host cells that contain the polynucleotide encoding KPP and that express KPP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

30 Immunological methods for detecting and measuring the expression of KPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive

35 binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al.

(1990) <u>Serological Methods, a Laboratory Manual</u>, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) <u>Current Protocols in Immunology</u>, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) <u>Immunochemical Protocols</u>, Humana Press, Totowa NJ).

5   A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding KPP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available,

10   and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as

15   well as substrates, cofactors, inhibitors, magnetic particles, and the like.

     Host cells transformed with polynucleotides encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing

20   polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

     In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation,

25   glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct

30   modification and processing of the foreign protein.

     In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the

35   screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide

moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion

5 proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so

10 that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled KPP may be achieved *in vitro* using the

15 TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that

20 specifically bind to KPP. One or more test compounds may be screened for specific binding to KPP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to KPP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of KPP can be used to screen for binding of test compounds,

25 such as antibodies, to KPP, a variant of KPP, or a combination of KPP and/or one or more variants KPP. In an embodiment, a variant of KPP can be used to screen for compounds that bind to a variant of KPP, but not to KPP having the exact sequence of a sequence of SEQ ID NO:1-5. KPP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to KPP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence

30 identity.

In an embodiment, a compound identified in a screen for specific binding to KPP can be closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can

35 be a natural ligand of a receptor KPP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140;

Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to KPP can be closely related to the natural receptor to which KPP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for KPP which is capable of propagating a signal, or a decoy receptor for KPP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of KPP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of KPP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of KPP.

In an embodiment, anticalins can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit KPP involves producing appropriate cells which express KPP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is

51

analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in

5 solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

10 . An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands .

15 (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

20 KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test

25 compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a

30 plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example,

35 mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and

52

grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using·the Cre-loxP

5   system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous

10 strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated *in vitro* in ES cells derived from human·blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al.

15 (1998) Science 282:1145-1147).

Polynucleotides encoding KPP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the

20 blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

25   · Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP can be found in Table 6 and can also be found in Example XI. Therefore, KPP appears to play a role in ·cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of

30 disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP -expression or activity, it is desirable to·increase the expression or activity·of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP.

35 Examples of such disorders include, but are not limited to, a cardiovascular disease such as

53

arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial
dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and
complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass
graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction,

5   hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis,
congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever
and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis
of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis,
neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation,

10   congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism,
pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis,
obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease,
emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and
mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases,

15   pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis,
hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia,
diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary
hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar
proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax,

20   pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung
transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS),
Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis,
anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune
polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

25   dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema,
episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic
gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's
thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,
myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis,

30   psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic
anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative
colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal
circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a
neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms;

35   Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other

extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous

5   system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic

10   nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses,

15   postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic

20   dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy,

25   spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and

30   ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous

35   xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease,

hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma,.

5   melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

10   In another embodiment, a vector capable of expressing KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent

15   a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those listed above.

20   In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may

25   be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

30   In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various

35   disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with

lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art. In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using

5   methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics,

10   and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with KPP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species,

15   various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

20   It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the

25   chimeric molecule may be produced.

Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol.

30   Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad.

35   Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985)

PF-1688 P

Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R.

5   (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

10   Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989)

15   Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific

20   antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an

25   association constant, K_a, which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a

30   particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody

35   (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell,

PF-1688 P

J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For

5   example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, supra; Coligan et al., supra).

10   In another embodiment of the invention, polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments

15   can be designed from various locations along the coding or control regions of sequences encoding KPP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence

20   complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278; Ausubel et al., supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene

25   delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency

30   (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

35   Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial

hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,

R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii)

express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated-

cell proliferation), or (iii) express a protein which affords protection against intracellular parasites

5   (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D.

(1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399),

hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides*

brasiliensis; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the

case where a genetic deficiency in KPP expression or regulation causes disease, the expression of

10   KPP from an appropriate population of transduced cells may alleviate the clinical manifestations

caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP

are treated by constructing mammalian expression vectors encoding KPP and introducing these

vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with

15   cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold

particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v)

the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-

217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol.

9:445-450).

20   Expression vectors that may be effective for the expression of KPP include, but are not

limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors

(Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA),

and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA). KPP

may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV),

25   Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible

promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl.

Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and

H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid

(Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND;

30   Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter

(Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of

the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver

35   polynucleotides to target cells in culture and require minimal effort to optimize experimental

parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

5   In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus cis-acting RNA sequences and coding sequences

10   required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.

15   (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by

20   reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4$^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-

25   2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be

30   versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

35   In another embodiment, a herpes-based, gene therapy delivery system is used to deliver

polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with

5 ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395).. The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which

10 consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of

15 recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus,

20 Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity

25 (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that

30 the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and

35 performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of

5   polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

10   Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding KPP.

15   Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of

20   candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively,

25   RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

30   modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine,

35   cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous

endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi

5 is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and

10 Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that

15 result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell.

20 Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3'

25 adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can

30 then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

35 In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in

selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to

5 target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods
10 can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray
15 methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not
20 limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP
25 expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in
30 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a
35 library of chemical compounds created combinatorially or randomly. A sample comprising a

polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is

5 detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the

10 polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a

15 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable

20 for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

25 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable

30 excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

In various embodiments, the compositions described herein, such as pharmaceutical

35 compositions, may be administered by any number of routes including, but not limited to, oral,

intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

.Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the

5   case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle

10   injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions, wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of

15   macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et

20   al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and

25   routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by

30   calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is

35   preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity.

The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding KPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect

68

and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides,

5 including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

10 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:6-10 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for polynucleotides encoding KPP include

15 the cloning of polynucleotides encoding KPP or KPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase

20 coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular

25 tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic

30 endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive

35 pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial

pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's

5   syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

10   allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

15   syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

20   complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis

25   pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis,

30   tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and

35   toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety,

70

and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis,

5   arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome,

10   myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine

15   palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal

20   hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as

25   actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract,

30   heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. Polynucleotides encoding KPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP

35   expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding KPP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization

5   complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered ·in comparison to a control sample then the presence of altered levels of polynucleotides encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials,

10 or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification.

15 Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

20 Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

25 With·respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may·provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development

30 or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding

35 KPP, and will be employed under optimized conditions for identification of a specific gene or

72

condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions
5 and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding KPP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids,
10 and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of
15 identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput
20 MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding
25 lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in
30 diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).
35 Methods which may also be used to quantify the expression of KPP include radiolabeling or

73

biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is

5   presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers

10   of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used

15   to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used

20   as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

25   quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

30   hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*,

35   as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

5   molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number

10   of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids

15   in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

20   In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with

25   levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected

30   individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric

35   focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate

slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of

5   equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its

10   partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for KPP to quantify the levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by contacting the microarray with the sample and

15   detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

20   Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which

25   alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated

30 · biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the

35   polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological

5 sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA

10 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

15 In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence.. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal

20 mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154).

25 Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic

30 map data (Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

35 In situ hybridization of chromosomal preparations and physical mapping techniques, such as

linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene

5   discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to

10   translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

15   between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound

20   KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In

25   this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

In additional embodiments, the nucleotide sequences which encode KPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

30   properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

35   The disclosures of all patents, applications and publications, mentioned above and below, are

expressly incorporated by reference herein.

EXAMPLES

I.   Construction of cDNA Libraries

5   Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte,

Palo Alto CA) and shown in Table 4, column 3. Some tissues are homogenized and lysed in

guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable

mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine

isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with

10   chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and

ethanol, or by other routine methods.

          Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity.

In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo

d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth

15   CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly

from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit

(Ambion, Austin TX).

          In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA

libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP

20.   vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended

procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is

initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double

stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For

most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE

25   CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative

agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the

polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid

(Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-

TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA),

30   pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed

into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α,

DH10B, or ElectroMAX DH10B from Invitrogen.

II.   Isolation of cDNA Clones

          Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision

35   using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least

one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml

5 of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using

10 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows. Sequencing reactions are processed using standard methods or high-throughput instrumentation such

15 as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic

20 separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the

25 cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such

30 as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res.

35 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc.

Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are

5 assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length

10 polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based

15 protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program

20 (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate

25 references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide

30 and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:6-10. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV.   Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases are initially identified by running the Genscan gene

35 identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is

a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a

5   FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides are analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases are also identified by homology to Incyte cDNA sequences that have been annotated as kinases and

10 . phosphatases. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then 'edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription.

15 When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

20   .V.   Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon

25 predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full-length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by

30 transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between

35 intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to

82

genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional

5   cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases

10   using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The

15   · GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous. genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI.   Chromosomal Mapping of KPP Encoding Polynucleotides

20   The sequences used to assemble SEQ ID NO:6-10 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:6-10 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as

25   the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map

30   position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation

35   hybrid markers whose sequences were included in each of the clusters. Human genome maps and

other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII.   Analysis of Polynucleotide Expression

5   Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related

10·   molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

15
$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is

20   calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated·by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate

25   the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79%

30   identity and 100% overlap.

Alternatively, polynucleotides encoding KPP are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the

35   following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ

cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas;
respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract.
The number of libraries in each category is counted and divided by the total number of libraries
across all categories. Similarly, each human tissue is classified into one of the following

5   disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma,
cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided
by the total number of libraries across all categories. The resulting percentages reflect the tissue- and
disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue
information are found in the LIFESEQ database (Incyte, Palo Alto CA).

10   **VIII.**   **Extension of KPP Encoding Polynucleotides**

       Full length polynucleotides are produced by extension of an appropriate fragment of the full
length molecule using oligonucleotide primers designed from this fragment. One primer is
synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to
initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06

15   software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in
length, to have a GC content of about 50% or more, and to anneal to the target sequence at
temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin
structures and primer-primer dimerizations is avoided.

       Selected human cDNA libraries are used to extend the sequence. If more than one extension

20   is necessary or desired, additional or nested sets of primers are designed.

       High fidelity amplification is obtained by PCR using methods well known in the art. PCR is
performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix
contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-
mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen),

25   and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI
B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps
2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the
parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec;
Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5

30   min; Step 7: storage at 4°C.

       The concentration of DNA in each well is determined by dispensing 100 µl PICOGREEN
quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE
and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,
Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II

35   (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and
5 sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun
. sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones are religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected
10 into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3
15 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC
20 DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

25 **IX.** · **Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides**
Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are. identified in SEQ ID NO:6-10 using the LIFESEQ database (Incyte). Sequences from the same gene—— are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish ·
30 SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify
35 errors introduced during laboratory processing, such as those caused by reverse transcriptase,

polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

5   Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The

10   Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

15   **X.   Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:6-10 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06

20   software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of

25   human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature

30   under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI.   Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

35   photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*),

mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a

5   procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat.

10   Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in

15   the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element

20   on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is

25   reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription

30   from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1

35   mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion

PF-1688 P

using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element

5   is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

10.   slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

15   Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

20   Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

25   Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the

30   addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

35   Reporter-labeled hybridization complexes are detected with a microscope equipped with an

89

Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-

5   scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores.

10   Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

15   The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different

20   fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC

25   computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

30   A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background

35   ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be

90

differentially expressed.

Expression

      For example, expression of SEQ ID NO:10 was down-regulated in diseased lung tissue versus normal lung tissue as determined by microarray analysis. Expression of SEQ ID NO:10 was

5   decreased at least two-fold in the lung tumor tissue with squamous cell carcinoma as compared to grossly uninvolved lung tissue from the same donor using a pair comparison experimental design. Therefore, in various embodiments, SEQ ID NO:10 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

10 **XII.   Complementary Polynucleotides**

      Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO

15 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

XIII.   Expression of KPP

20   Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

25 Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP

30 by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA

35 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on

5   immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites._FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins

10   (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX, and XXI, where applicable.

XIV.   Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically

15   elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either

20   liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated,

25   laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side

30   light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

35   The influence of KPP on gene expression can be assessed using highly purified populations

of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success

5   NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV.   Production of KPP Specific Antibodies

KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,

10   Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for

15   selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

20   increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI.   Purification of Naturally Occurring KPP Using Specific Antibodies

25   Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

30   Media containing KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

35   ## XVII.   Identification of Molecules Which Interact with KPP

KPP, or biologically active fragments thereof, are labeled with [125]I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used

5 to calculate values for the number, affinity, and association of KPP with the candidate molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)

10 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein

15 substrate by KPP in the presence of $[\gamma-^{32}P]ATP$. KPP is incubated with the protein substrate, ^{32}P-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P-ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

20 In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P-ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P-peptide product. The binding sample then undergoes a centrifugal ultrafiltration

25 process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P-ATP. The reservoir of the centrifuged unit containing the ^{32}P-peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation,

30 Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and $p34^{cdc2}$kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and src kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in an assay containing

35 KPP, 50 μl of kinase buffer, 1 μg substrate, such as myelin basic protein (MBP) or synthetic peptide

substrates, 1 mM DTT, 10 µg ATP, and 0.5 µCi [γ-^{32}P]ATP. The reaction is incubated at 30°C for 30·
minutes and stopped by pipetting onto P81 paper. The unincorporated [γ-^{32}P]ATP is removed by
washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively,
the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a

5   12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated ^{32}P is
proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be
measured by the incorporation of ^{32}P from [γ-^{32}P]ATP into ADP or GDP using a gamma radioisotope
counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide

10   mono-phosphate substrate (AMP or GMP) and ^{32}P-labeled ATP as the phosphate donor. The reaction
is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is
neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide
fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is
proportional to the activity of KPP.

15   ·   In yet another alternative, other assays for KPP include scintillation proximity assays (SPA),
scintillation plate technology and filter binding assays. Useful substrates include recombinant
·proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin.
Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by
such assays.

20   In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-
nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the
presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6
ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid
phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 µl of 10

·25   mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 µl of 40 mM NaCl at 37°C for 20 min. The
reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997)
J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the _ _ _
hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is
proportional to the activity of KPP in the assay.

30   In the alternative, KPP activity is determined by measuring the amount of phosphate removed
from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final
volume of 30 µl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1%
β-mercaptoethanol and 10 µM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate.
Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched

35   with 450 µl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM Na$_4$P$_2$O$_7$, and 2 mM NaH$_2$PO$_4$, then

95

centrifuged at $12,000 \times g$ for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP

5 with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of ^{125}I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter

10 (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated ^{32}P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. KPP

15 activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance KPP activity.

XXI. Identification of KPP Substrates

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may

20 be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in

25 *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by. – — chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated

30 transfection with 20 μg of CsCl-purified DNA per 10-cm dish of cells or 8 μg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris-HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μg/ml leupeptin/5 μg/ml aprotinin/1 mM benzamidine

35 (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an

PF-1688 P

appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μg of mAb or 10 μl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

5

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for

10   using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined

15   with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

PF-1688 P

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

    a)   a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5,

    b)   a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-2,

    c)   a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to the amino acid sequence of SEQ ID NO:5,

    d)   a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and .

    e)   an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

    a)   culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

98

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b)   recovering the polypeptide so expressed.

5   10.· A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

10   12. An isolated polynucleotide selected from the group consisting of:

a)   a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10,

b)   a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-8,

15

c)   a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to the polynucleotide sequence of SEQ ID NO:10,

d)   a polynucleotide complementary to a polynucleotide of a),

e)   a polynucleotide complementary to a polynucleotide of b),

20   f)   a polynucleotide complementary to a polynucleotide of c), and

g)   an RNA equivalent of a)-f).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

25

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a)   hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample,

30   and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said · target polynucleotide·or fragments thereof, and

b)   detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

35

99

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

5   a)   amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

     b)   detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

10   17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

15

19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of claim 17.

20   20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

     a)   contacting a sample comprising a polypeptide of claim 1 with a compound, and

     b)   detecting agonist activity in the sample.

25   21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of 30 claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

     a)   contacting a sample comprising a polypeptide of claim 1 with a compound, and

35   b)   detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional
5   KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

   a)   combining the polypeptide of claim 1 with at least one test compound under suitable
10   conditions, and

   b)   detecting binding of the polypeptide of claim 1 to the test compound, thereby
         identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of
15   claim 1, the method comprising:

   a)   combining the polypeptide of claim 1 with at least one test compound under
         conditions permissive for the activity of the polypeptide of claim 1,

   b)   assessing the activity of the polypeptide of claim 1 in the presence of the test
         compound, and

20 c)   comparing the activity of the polypeptide of claim 1 in the presence of the test
         compound with the activity of the polypeptide of claim 1 in the absence of the test
         compound, wherein a change in the activity of the polypeptide of claim 1 in the
         presence of the test compound is indicative of a compound that modulates the activity
         of the polypeptide of claim 1.
25

28. A method of screening a compound for effectiveness in altering expression of a target
polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method
comprising:

   a)   contacting a sample comprising the target polynucleotide with a compound, under
30   conditions suitable for the expression of the target polynucleotide,

   b)   detecting altered expression of the target polynucleotide, and

   c)   comparing the expression of the target polynucleotide in the presence of varying
         amounts of the compound and in the absence of the compound.

35   29. A method of screening for potential toxicity of a test compound, the method comprising:

PF-1688 P

a)   treating a biological sample containing nucleic acids with the test compound,

b)   hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

5   polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c)   quantifying the amount of hybridization complex, and

d)   comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a

10   difference in the amount of hybridization complex in the treated biological sample indicates potential toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of KPP in a biological sample, the method comprising:

15   a)   combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b)   detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

20

31. The antibody of claim 11, wherein the antibody is:

a)   a chimeric antibody,

b)   a single chain antibody,

c)   a Fab fragment,

25   d)   a F(ab')$_2$ fragment, or

e)   a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

30   33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35

35. A method of diagnosing a condition or disease associated with the expression of KPP in

a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a)   immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b)   isolating antibodies from the animal, and

c)   screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a)   immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b)   isolating antibody producing cells from the animal,

c)   fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d)   culturing the hybridoma cells, and

e)   isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 in a sample, the method comprising:

    a)   incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

    b)   detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 from a sample, the method comprising:

    a)   incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

    b)   separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

    a)   labeling the polynucleotides of the sample,

    b)   contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

    c)   quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:6.

62. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:7.

63. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:8.

64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.

5   65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.

ABSTRACT OF THE DISCLOSURE

Various embodiments of the invention provide human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. Embodiments of the invention also provide

5   expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

PF-1688 P

<110> YUE, Henry; NGUYEN, Danniel B.
      THORNTON, Michael B.; GURURAJAN, Rajagopal
      GANDHI, Ameena R.; LU, Yan
      YAO, Monique G.; BAUGHN, Mariah R.
      CHAWLA, Narinder; LI, Joana X.
      LUO, Wen; LEE, Ernestine A.
      FORSYTHE, Ian J.; BECHA, Shanya D.
      ISON, Craig H.

<120> KINASES AND PHOSPHATASES

<130> PF-1688 P

<140> To Be Assigned
<141> Herewith

<160> 10
<170> PERL Program

<210> 1
<211> 1125
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2509577CD1

<400> 1
Met Pro Asp Gln Asp Lys Lys Val Lys Thr Thr Glu Lys Ser Thr
1   5   10   15
Asp Lys Gln Gln Glu Ile Thr Ile Arg Asp Tyr Ser Asp Leu Lys
                20   25   30
Arg Leu Arg Cys Leu Leu Asn Val Gln Ser Ser Lys Gln Gln Leu
                35   40   45
Pro Ala Ile Asn Phe Asp Ser Ala Gln Asn Ser Met Thr Lys Ser
                50   55   60
Glu Pro Ala Ile Arg Ala Gly Gly His Arg Ala Arg Gly Gln Trp
                65   70   75
His Glu Ser Thr Glu Ala Val Glu Leu Glu Asn Phe Ser Ile Asn
                80   85   90
Tyr Lys Asn Glu Arg Asn Phe Ser Lys His Pro Gln Arg Lys Leu
                95   100   105
Phe Gln Glu Ile Phe Thr Ala Leu Val Lys Asn Arg Leu Ile Ser
                110   115   120
Arg Glu Trp Val Asn Arg Ala Pro Ser Ile His Phe Leu Arg Val
                125   130   135
Leu Ile Cys Leu Arg Leu Leu Met Arg Asp Pro Cys Tyr Gln Glu
                140   145   150
Ile Leu His Ser Leu Gly Gly Ile Glu Asn Leu Ala Gln Tyr Met
                155   160   165
Glu Ile Val Ala Asn Glu Tyr Leu Gly Tyr Gly Glu Glu Gln His
                170   175   180
Thr Val Asp Lys Leu Val Asn Met Thr Tyr Ile Phe Gln Lys Leu
                185   190   195
Ala Ala Val Lys Asp Gln Arg Glu Trp Val Thr Thr Ser Gly Ala
                200   205   210
His Lys Thr Leu Val Asn Leu Leu Gly Ala Arg Asp Thr Asn Val
                215   220   225

1

PF-1688 P

```
Leu Leu Gly Ser Leu Leu Ala Leu Ala Ser Leu Ala Glu Ser Gln
            230               235                   240
Glu Cys Arg Glu Lys Ile Ser Glu Leu Asn Ile Val Glu Asn Leu
            245               250                   255
Leu Met Ile Leu His Glu Tyr Asp Leu Leu Ser Lys Arg Leu Thr
            260               265                   270
Ala Glu Leu Leu Arg Leu Leu Cys Ala Glu Pro Gln Val Lys Glu
            275               280                   285
Gln Val Lys Leu Tyr Glu Gly Ile Pro Val Leu Leu Ser Leu Leu
            290               295                   300
His Ser Asp His Leu Lys Leu Leu Trp Ser Ile Val Trp Ile Leu
            305               310                   315
Val Gln Val Cys Glu Asp Pro Glu Thr Ser Val Glu Ile Arg Ile
            320               325                   330
Trp Gly Gly Ile Lys Gln Leu Leu His Ile Leu Gln Gly Asp Arg
            335               340                   345
Asn Phe Val Ser Asp His Ser Ser Ile Gly Ser Leu Ser Ser Ala
            350               355                   360
Asn Ala Ala Gly Arg Ile Gln Gln Leu His Leu Ser Glu Asp Leu
            365               370                   375
Ser Pro Arg Glu Ile Gln Glu Asn Thr Phe Ser Leu Gln Ala Ala
            380               385                   390
Cys Cys Ala Ala Leu Thr Glu Leu Val Leu Asn Asp Thr Asn Ala
            395               400                   405
His Gln Val Val Gln Glu Asn Gly Val Tyr Thr Ile Ala Lys Leu
            410               415                   420
Ile Leu Pro Asn Lys Gln Lys Asn Ala Ala Lys Ser Asn Leu Leu
            425               430                   435
Gln Cys Tyr Ala Phe Arg Ala Leu Arg Phe Leu Phe Ser Met Glu
            440               445                   450
Arg Asn Arg Pro Leu Phe Lys Arg Leu Phe Pro Thr Asp Leu Phe
            455               460                   465
Glu Ile Phe Ile Asp Ile Gly His Tyr Val Arg Asp Ile Ser Ala
            470               475                   480
Tyr Glu Glu Leu Val Ser Lys Leu Asn Leu Leu Val Glu Asp Glu
            485               490                   495
Leu Lys Gln Ile Ala Glu Asn Ile Glu Ser Ile Asn Gln Asn Lys
            500               505                   510
Ala Pro Leu Lys Tyr Ile Gly Asn Tyr Ala Ile Leu Asp His Leu
            515               520                   525
Gly Ser Gly Ala Phe Gly Cys Val Tyr Lys Val Arg Lys His Ser
            530               535                   540
Gly Gln Asn Leu Leu Ala Met Lys Glu Val Asn Leu His Asn Pro
            545               550                   555
Ala Phe Gly Lys Asp Lys Lys Asp Arg Asp Ser Ser Val Arg Asn
            560               565                   570
Ile Val Ser Glu Leu Thr Ile Ile Lys Glu Gln Leu Tyr His Pro
            575               580                   585
Asn Ile Val Arg Tyr Tyr Lys Thr Phe Leu Glu Asn Asp Arg Leu
            590               595                   600
Tyr Ile Val Met Glu Leu Ile Glu Gly Ala Pro Leu Gly Glu His
            605               610                   615
Phe Ser Ser Leu Lys Glu Lys His His His Phe Thr Glu Glu Arg
            620               625                   630
Leu Trp Lys Ile Phe Ile Gln Leu Cys Leu Ala Leu Arg Tyr Leu
            635               640                   645
His Lys Glu Lys Arg Ile Val His Arg Asp Leu Thr Pro Asn Asn
            650               655                   660
Ile Met Leu Gly Asp Lys Asp Lys Val Thr Val Thr Asp Phe Gly
```

2

PF-1688 P

```
              665                 670                 675
Leu Ala Lys Gln Lys Gln Glu Asn Ser Lys Leu Thr Ser Val Val
              680                 685                 690
Gly Thr Ile Leu Tyr Ser Cys Pro Glu Val Leu Lys Ser Glu Pro
              695                 700                 705
Tyr Gly Glu Lys Ala Asp Val Trp Ala Val Gly Cys Ile Leu Tyr
              710                 715                 720
Gln Met Ala Thr Leu Ser Pro Pro Phe Tyr Ser Thr Asn Met Leu
              725                 730                 735
Ser Leu Ala Thr Lys Ile Val Glu Ala Val Tyr Glu Pro Val Pro
              740                 745                 750
Glu Gly Ile Tyr Ser Glu Lys Val Thr Asp Thr Ile Ser Arg Cys
              755                 760                 765
Leu Thr Pro Asp Ala Glu Ala Arg Pro Asp Ile Val Glu Val Ser
              770                 775                 780
Ser Met Ile Ser Asp Val Met Met Lys Tyr Leu Asp Asn Leu Ser
              785                 790                 795
Thr Ser Gln Leu Ser Leu Glu Lys Lys Leu Glu Arg Glu Arg Arg
              800                 805                 810
Arg Thr Gln Arg Tyr Phe Met Glu Ala Asn Arg Asn Thr Val Thr
              815                 820                 825
Cys His His Glu Leu Ala Val Leu Ser His Glu Thr Phe Glu Lys
              830                 835                 840
Ala Ser Leu Ser Ser Ser Ser Ser Gly Ala Ala Ser Leu Lys Ser
              845                 850                 855
Glu Leu Ser Glu Ser Ala Asp Leu Pro Pro Glu Gly Phe Gln Ala
              860                 865                 870
Ser Tyr Gly Lys Asp Glu Asp Arg Ala Cys Asp Glu Ile Leu Ser
              875                 880                 885
Asp Asp Asn Phe Asn Leu Glu Asn Ala Glu Lys Asp Thr Tyr Ser
              890                 895                 900
Glu Val Asp Asp Glu Leu Asp Ile Ser Asp Asn Ser Ser Ser Ser
              905                 910                 915
Ser Ser Ser Pro Leu Lys Glu Ser Thr Phe Asn Ile Leu Lys Arg
              920                 925                 930
Ser Phe Ser Ala Ser Gly Gly Glu Arg Gln Ser Gln Thr Arg Asp
              935                 940                 945
Phe Thr Gly Gly Thr Gly Ser Arg Pro Arg Pro Gly Pro Gln Met
              950                 955                 960
Gly Thr Phe Leu Trp Gln Ala Ser Ala Gly Ile Ala Val Ser Gln
              965                 970                 975
Arg Lys Val Arg Gln Ile Ser Asp Pro Ile Gln Gln Ile Leu Ile
              980                 985                 990
Gln Leu His Lys Ile Ile Tyr Ile Thr Gln Leu Pro Pro Ala Leu
              995                 1000                1005
His His Asn Leu Lys Arg Arg Val Ile Glu Arg Phe Lys Lys Ser
              1010                1015                1020
Leu Phe Ser Gln Gln Ser Asn Pro Cys Asn Leu Lys Ser Glu Ile
              1025                1030                1035
Lys Lys Leu Ser Gln Gly Ser Pro Glu Pro Ile Glu Pro Asn Phe
              1040                1045                1050
Phe Thr Ala Asp Tyr His Leu Leu His Arg Ser Ser Gly Gly Asn
              1055                1060                1065
Ser Leu Ser Pro Asn Asp Pro Thr Gly Leu Pro Thr Ser Ile Glu
              1070                1075                1080
Leu Glu Glu Gly Ile Thr Tyr Glu Gln Met Gln Thr Val Ile Glu
              1085                1090                1095
Glu Val Leu Glu Glu Ser Gly Tyr Tyr Asn Phe Thr Ser Asn Arg
              1100                1105                1110
```

3

PF-1688 P

Tyr His Ser Tyr Pro Trp Gly Thr Lys Asn His Pro Thr Lys Arg
            1115   1120   1125

<210> 2
<211> 888
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7505222CD1

<400> 2
Met Gln Ile Val Gly Ser Pro Gly Pro Gly Ala Ala Trp Pro Val
 1   5   10   15
Lys Arg Val Val Phe Pro Asn Gly Glu Gln Phe Leu Leu Ser Val
                20   25   30
Ala Thr Lys Lys Val Ile Cys Leu Cys Leu Gly Lys Ala Gly Arg
                35   40   45
Lys Val Leu Ala Lys Lys Leu Ser Pro Leu Glu Thr Met Asp Lys
                50   55   60
Tyr Asp Val Ile Lys Ala Ile Gly Gln Gly Ala Phe Gly Lys Ala
                65   70   75
Tyr Leu Ala Lys Gly Lys Ser Asp Ser Lys His Cys Val Ile Lys
                80   85   90
Glu Ile Asn Phe Glu Lys Met Pro Ile Gln Glu Lys Glu Ala Ser
                95   100   105
Lys Lys Glu Val Ile Leu Leu Glu Lys Met Lys His Pro Asn Ile
                110   115   120
Val Ala Phe Phe Asn Ser Phe Gln Glu Asn Gly Arg Leu Phe Ile
                125   130   135
Val Met Glu Tyr Cys Asp Gly Gly Asp Leu Met Lys Arg Ile Asn
                140   145   150
Arg Gln Arg Gly Val Leu Phe Ser Glu Asp Gln Ile Leu Gly Trp
                155   160   165
Phe Val Gln Ile Ser Leu Gly Leu Lys His Ile His Asp Arg Lys
                170   175   180
Ile Leu His Arg Asp Ile Lys Ala Gln Asn Ile Phe Leu Ser Lys
                185   190   195
Asn Gly Met Val Ala Lys Leu Gly Asp Phe Gly Ile Ala Arg Val
                200   205   210
Leu Asn Asn Ser Met Glu Leu Ala Arg Thr Cys Ile Gly Thr Pro
                215   220   225
Tyr Tyr Leu Ser Pro Glu Ile Cys Gln Asn Lys Pro Tyr Asn Asn
                230   235   240
Lys Thr Asp Ile Trp Ser Leu Gly Cys Val Leu Tyr Glu Leu Cys
                245   250   255
Thr Leu Lys His Pro Phe Glu Gly Asn Asn Leu Gln Gln Leu Val
                260   265   270
Leu Lys Ile Cys Gln Ala His Phe Ala Pro Ile Ser Pro Gly Phe
                275   280   285
Ser Arg Glu Leu His Ser Leu Ile Ser Gln Leu Phe Gln Val Ser
                290   295   300
Pro Arg Asp Arg Pro Ser Ile Asn Ser Ile Leu Lys Arg Pro Phe
                305   310   315
Leu Glu Asn Leu Ile Pro Lys Tyr Leu Thr Pro Glu Val Ile Gln
                320   325   330
Glu Glu Phe Ser His Met Leu Ile Cys Arg Ala Gly Ala Pro Ala

4

PF-1688 P

		335					340				345
Ser	Arg	His	Ala	Gly	Lys	Val	Val	Gln	Lys	Cys	Lys	Ile	Gln	Lys

```
              335                          340                          345
Ser Arg His Ala Gly Lys Val Val Gln Lys Cys Lys Ile Gln Lys
              350                          355                          360
Val Arg Phe Gln Gly Lys Cys Pro Pro Arg Ser Arg Ile Ser Val
              365                          370                          375
Pro Ile Lys Arg Asn Ala Ile Leu His Arg Asn Glu Trp Arg Pro
              380                          385                          390
Pro Ala Gly Ala Gln Lys Ala Arg Ser Ile Lys Met Ile Glu Arg
              395                          400                          405
Pro Lys Ile Ala Ala Val Cys Gly His Tyr Asp Tyr Tyr Tyr Ala
              410                          415                          420
Gln Leu Asp Met Leu Arg Arg Arg Ala His Lys Pro Ser Tyr His
              425                          430                          435
Pro Ile Pro Gln Glu Asn Thr Gly Val Glu Asp Tyr Gly Gln Glu
              440                          445                          450
Thr Arg His Gly Pro Ser Pro Ser Gln Trp Pro Ala Glu Tyr Leu
              455                          460                          465
Gln Arg Lys Phe Glu Ala Gln Gln Tyr Lys Leu Lys Val Glu Lys
              470                          475                          480
Gln Leu Gly Leu Arg Pro Ser Ser Ala Glu Pro Asn Tyr Asn Gln
              485                          490                          495
Arg Gln Glu Leu Arg Ser Asn Gly Glu Glu Pro Arg Phe Gln Glu
              500                          505                          510
Leu Pro Phe Arg Lys Asn Glu Met Lys Glu Gln Glu Tyr Trp Lys
              515                          520                          525
Gln Leu Glu Glu Ile Arg Gln Gln Tyr His Asn Asp Met Lys Glu
              530                          535                          540
Ile Arg Lys Lys Met Gly Arg Glu Pro Glu Glu Asn Ser Lys Ile
              545                          550                          555
Ser His Lys Thr Tyr Leu Val Lys Lys Ser Asn Leu Pro Val His
              560                          565                          570
Gln Asp Ala Ser Glu Gly Glu Ala Pro Val Gln Asp Ile Glu Lys
              575                          580                          585
Asp Leu Lys Gln Met Arg Leu Gln Asn Thr Lys Glu Ser Lys Asn
              590                          595                          600
Pro Glu Gln Lys Tyr Lys Ala Lys Gly Val Lys Phe Glu Ile Asn
              605                          610                          615
Leu Asp Lys Cys Ile Ser Asp Glu Asn Ile Leu Gln Glu Glu Glu
              620                          625                          630
Ala Met Asp Ile Pro Asn Glu Thr Leu Thr Phe Glu Asp Gly Met
              635                          640                          645
Lys Phe Lys Glu Tyr Glu Cys Val Lys Glu His Gly Asp Tyr Thr
              650                          655                          660
Asp Lys Ala Phe Glu Lys Leu His Cys Pro Glu Ala Gly Phe Ser
              665                          670                          675
Thr Gln Thr Val Ala Ala Val Gly Asn Arg Arg Gln Trp Asp Gly
              680                          685                          690
Gly Ala Pro Gln Thr Leu Leu Gln Met Met Ala Val Ala Asp Ile
              695                          700                          705
Thr Ser Thr Cys Pro Thr Gly Pro Asp Asn Gly Gln Val Ile Val
              710                          715                          720
Ile Glu Gly Ile Pro Gly Asn Arg Lys Gln Trp Arg His Glu Ala
              725                          730                          735
Pro Gly Thr Leu Met Ser Val Leu Ala Ala Ala His Leu Thr Ser
              740                          745                          750
Ser Ser Phe Ser Ala Asp Glu Glu Phe Ala Met Gly Thr Leu Lys
              755                          760                          765
Gln Trp Leu Pro Lys Glu Glu Asp Glu Gly Lys Val Glu Met Val
              770                          775                          780
```

5

PF-1688 P

```
Ser Gly Ile Glu Val Asp Glu Glu Gln Leu Glu Pro Arg Ser Asp
              785                   790                   795
Asp Asp Asp Thr Asn Phe Glu Glu Ser Glu Asp Glu Leu Arg Asp
              800                   805                   810
Glu Val Val Glu Tyr Leu Glu Lys Leu Ala Thr Phe Lys Gly Glu
              815                   820                   825
Glu Lys Thr Glu Glu Ala Ser Ser Thr Ser Lys Asp Ser Arg Lys
              830                   835                   840
Ser Arg Glu Arg Glu Gly Ile Ser Met Gln Lys Ser Glu Glu Leu
              845                   850                   855
Arg Glu Gly Leu Glu Asn Ile Ser Thr Thr Ser Asn Asp His Ile
              860                   865                   870
Cys Ile Thr Asp Glu Asp Gln Gly Thr Ser Thr Thr Ser Gln Asn
              875                   880                   885
Ile Gln Val
```

<210> 3
<211> 487
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524408CD1

<400> 3
```
Met Gly Arg Ile Gly Ile Ser Cys Leu Phe Pro Ala Ser Trp His
  1               5                  10                   15
Phe Ser Ile Ser Pro Val Gly Cys Pro Arg Ile Leu Asn Thr Asn
              20                   25                    30
Leu Arg Gln Ile Met Val Ile Ser Val Leu Ala Ala Ala Val Ser
              35                    40                    45
Leu Leu Tyr Phe Ser Val Val Ile Ile Arg Asn Lys Tyr Gly Arg
              50                    55                    60
Leu Thr Arg Asp Lys Lys Phe Gln Arg Tyr Leu Ala Arg Val Thr
              65                    70                    75
Asp Ile Glu Ala Thr Asp Thr Asn Asn Pro Asn Val Asn Tyr Gly
              80                    85                    90
Ile Val Val Asp Cys Gly Ser Ser Gly Ser Arg Val Phe Val Tyr
              95                   100                   105
Cys Trp Pro Arg His Asn Gly Asn Pro His Asp Leu Leu Asp Ile
              110                   115                   120
Arg Gln Met Arg Asp Lys Asn Arg Lys Pro Val Val Met Lys Ile
              125                   130                   135
Lys Pro Gly Ile Ser Glu Phe Ala Thr Ser Pro Glu Lys Val Ser
              140                   145                   150
Asp Tyr Ile Ser Pro Leu Leu Asn Phe Ala Ala Glu His Val Pro
              155                   160                   165
Arg Ala Lys His Lys Glu Thr Pro Leu Tyr Ile Leu Cys Thr Ala
              170                   175                   180
Gly Met Arg Ile Leu Pro Glu Ser Gln Gln Lys Ala Ile Leu Glu
              185                   190                   195
Asp Leu Leu Thr Asp Ile Pro Val His Phe Asp Phe Leu Phe Ser
              200                   205                   210
Asp Ser His Ala Glu Val Ile Ser Gly Lys Gln Glu Gly Val Tyr
              215                   220                   225
Ala Trp Ile Gly Ile Asn Phe Val Leu Gly Arg Phe Glu His Ile
              230                   235                   240
```

6

PF-1688 P

```
Glu Asp Asp Asp Glu Ala Val Val Glu Val Asn Ile Pro Gly Ser
            245                 250                 255
Val Ser Ser Glu Ala Ile Val Arg Lys Arg Thr Ala Gly Ile Leu
            260                 265                 270
Asp Met Gly Gly Val Leu Thr Gln Ile Ala Tyr Glu Val Pro Lys
            275                 280                 285
Thr Ala Ser Phe Ala Ser Ser Gln Gln Glu Glu Val Ala Lys Asn
            290                 295                 300
Leu Leu Ala Glu Phe Asn Leu Gly Cys Asp Val His Gln Thr Glu
            305                 310                 315
His Val Tyr Arg Val Tyr Val Ala Thr Phe Phe Gly Phe Gly Gly
            320                 325                 330
Asn Ala Ala Arg Gln Arg Tyr Glu Asp Arg Ile Phe Ala Asn Thr
            335                 340                 345
Ile Gln Lys Asn Arg Leu Leu Gly Lys Gln Thr Gly Leu Thr Pro
            350                 355                 360
Asp Met Pro Tyr Leu Asp Pro Cys Leu Pro Leu Asp Ile Lys Asp
            365                 370                 375
Glu Ile Gln Gln Asn Gly Gln Thr Ile Tyr Leu Arg Gly Thr Gly
            380                 385                 390
Asp Phe Asp Leu Cys Arg Glu Thr Ile Gln Pro Phe Met Asn Lys
            395                 400                 405
Thr Asn Glu Thr Gln Thr Ser Leu Asn Gly Val Tyr Gln Pro Pro
            410                 415                 420
Ile His Phe Gln Asn Ser Glu Phe Tyr Gly Phe Ser Glu Phe Tyr
            425                 430                 435
Tyr Cys Thr Glu Asp Val Leu Arg Met Gly Gly Asp Tyr Asn Ala
            440                 445                 450
Ala Lys Phe Thr Lys Ala Ala Lys Asp Tyr Cys Ala Thr Lys Trp
            455                 460                 465
Ser Ile Leu Arg Glu Arg Phe Asp Arg Gly Leu Tyr Ala Ser His
            470                 475                 480
Ala Asp Leu His Arg Leu Lys
            485
```

<210> 4
<211> 1309
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526163CD1

<400> 4
```
Met Asp Glu Ser Ser Leu Leu Arg Arg Arg Gly Leu Gln Lys Glu
1                 5                 10                  15
Leu Ser Leu Pro Arg Arg Gly Arg Gly Cys Arg Ser Gly Asn Arg
            20                  25                  30
Lys Ser Leu Val Val Gly Thr Pro Ser Pro Thr Leu Ser Arg Pro
            35                  40                  45
Leu Ser Pro Leu Ser Val Pro Thr Ala Gly Ser Ser Pro Leu Asp
            50                  55                  60
Ser Pro Arg Asn Phe Ser Ala Ala Ser Ala Leu Asn Phe Pro Phe
            65                  70                  75
Ala Arg Arg Ala Asp Gly Arg Arg Trp Ser Leu Ala Ser Leu Pro
            80                  85                  90
Ser Ser Gly Tyr Gly Thr Asn Thr Pro Ser Ser Thr Leu Ser Ser
            95                  100                 105
```

7

PF-1688 P

```
Ser Ser Ser Ser Arg Glu Arg Leu His Gln Leu Pro Phe Gln Pro
            110                 115                 120
Thr Pro Asp Glu Leu His Phe Leu Ser Lys His Phe Arg Ser Ser
            125                 130                 135
Glu Asn Val Leu Asp Glu Glu Gly Gly Arg Ser Pro Arg Leu Arg
            140                 145                 150
Pro Arg Ser Arg Ser Leu Ser Pro Gly Arg Ala Thr Gly Thr Phe
            155                 160                 165
Asp Asn Glu Ile Val Met Met Asn His Val Tyr Arg Glu Arg Phe
            170                 175                 180
Pro Lys Ala Thr Ala Gln Met Glu Gly Arg Leu Gln Glu Phe Leu
            185                 190                 195
Thr Ala Tyr Ala Pro Gly Ala Arg Leu Ala Leu Ala Asp Gly Val
            200                 205                 210
Leu Gly Phe Ile His His Gln Ile Val Glu Leu Ala Arg Asp Cys
            215                 220                 225
Leu Ala Lys Ser Gly Glu Asn Leu Val Thr Ser Arg Tyr Phe Leu
            230                 235                 240
Glu Met Gln Glu Lys Leu Glu Arg Leu Leu Gln Asp Ala His Glu
            245                 250                 255
Arg Ser Asp Ser Glu Glu Val Ser Phe Ile Val Gln Leu Val Arg
            260                 265                 270
Lys Leu Leu Ile Ile Ile Ser Arg Pro Ala Arg Leu Leu Glu Cys
            275                 280                 285
Leu Glu Phe Asp Pro Glu Glu Phe Tyr His Leu Leu Glu Ala Ala
            290                 295                 300
Glu Gly His Ala Arg Glu Gly Gln Gly Ile Lys Thr Asp Leu Pro
            305                 310                 315
Gln Tyr Ile Ile Gly Gln Leu Gly Leu Ala Lys Asp Pro Leu Glu
            320                 325                 330
Glu Met Val Pro Leu Ser His Leu Glu Glu Glu Gln Pro Pro Ala
            335                 340                 345
Pro Glu Ser Pro Glu Ser Arg Ala Leu Val Gly Gln Ser Arg Arg
            350                 355                 360
Lys Pro Cys Glu Ser Asp Phe Glu Thr Ile Lys Leu Ile Ser Asn
            365                 370                 375
Gly Ala Tyr Gly Ala Val Tyr Leu Val Arg His Arg Asp Thr Arg
            380                 385                 390
Gln Arg Phe Ala Ile Lys Lys Ile Asn Lys Gln Asn Leu Ile Leu
            395                 400                 405
Arg Asn Gln Val Gln Gln Val Phe Val Glu Arg Asp Ile Leu Thr
            410                 415                 420
Phe Ala Glu Asn Pro Phe Val Val Ser Met Phe Cys Ser Phe Glu
            425                 430                 435
Thr Arg Arg His Leu Cys Met Val Met Glu Tyr Val Glu Gly Gly
            440                 445                 450
Asp Cys Ala Thr Leu Leu Lys Asn Met Gly Pro Leu Pro Val Asp
            455                 460                 465
Met Ala Arg Leu Tyr Phe Ala Glu Thr Val Leu Ala Leu Glu Tyr
            470                 475                 480
Leu His Asn Tyr Gly Ile Val His Arg Asp Leu Lys Pro Asp Asn
            485                 490                 495
Leu Leu Ile Thr Ser Leu Gly His Ile Lys Leu Thr Asp Phe Gly
            500                 505                 510
Leu Ser Lys Ile Gly Leu Met Ser Met Ala Thr Asn Leu Tyr Glu
            515                 520                 525
Gly His Ile Glu Lys Asp Ala Arg Glu Phe Ile Asp Lys Gln Val
            530                 535                 540
Cys Gly Thr Pro Glu Tyr Ile Ala Pro Glu Val Ile Phe Arg Gln
```

8

PF-1688 P

```
                   545                      550                      555
    Gly Tyr Gly Lys Pro Val Asp Trp Trp Ala Met Gly Val Val Leu
                   560                      565                      570
    Tyr Glu Phe Leu Val Gly Cys Val Pro Phe Phe Gly Asp Thr Pro
                   575                      580                      585
    Glu Glu Leu Phe Gly Gln Val Val Ser Asp Glu Ile Met Trp Pro
                   590                      595                      600
    Glu Gly Asp Glu Ala Leu Pro Ala Asp Ala Gln Asp Leu Ile Thr
                   605                      610                      615
    Arg Leu Leu Arg Gln Ser Pro Leu Asp Arg Leu Gly Thr Gly Gly
                   620                      625                      630
    Thr His Glu Val Lys Gln His Pro Phe Phe Leu Ala Leu Asp Trp
                   635                      640                      645
    Ala Gly Leu Leu Arg His Lys Ala Glu Phe Val Pro Gln Leu Glu
                   650                      655                      660
    Ala Glu Asp Asp Thr Ser Tyr Phe Asp Thr Arg Ser Glu Arg Tyr
                   665                      670                      675
    Arg His Leu Gly Ser Glu Asp Asp Glu Thr Asn Asp Glu Glu Ser
                   680                      685                      690
    Ser Thr Glu Ile Pro Gln Phe Ser Ser Cys Ser His Arg Phe Ser
                   695                      700                      705
    Lys Val Tyr Ser Ser Ser Glu Phe Leu Ala Val Gln Pro Thr Pro
                   710                      715                      720
    Thr Phe Ala Glu Arg Ser Phe Ser Glu Asp Arg Glu Glu Gly Trp
                   725                      730                      735
    Glu Arg Ser Glu Val Asp Tyr Gly Arg Arg Leu Ser Ala Asp Ile
                   740                      745                      750
    Arg Leu Arg Ser Trp Thr Ser Ser Gly Ser Ser Cys Gln Ser Ser
                   755                      760                      765
    Ser Ser Gln Pro Glu Arg Gly Pro Ser Pro Ser Leu Leu Asn Thr
                   770                      775                      780
    Ile Ser Leu Asp Thr Met Pro Lys Phe Ala Phe Ser Ser Glu Asp
                   785                      790                      795
    Glu Gly Val Gly Pro Gly Pro Ala Gly Pro Lys Arg Pro Val Phe
                   800                      805                      810
    Ile Leu Gly Glu Pro Asp Pro Pro Ala Ala Thr Pro Val Met
                   815                      820                      825
    Pro Lys Pro Ser Ser Leu Ser Ala Asp Thr Ala Ala Leu Ser His
                   830                      835                      840
    Ala Arg Leu Arg Ser Asn Ser Ile Gly Ala Arg His Ser Thr Pro
                   845                      850                      855
    Arg Pro Leu Asp Ala Gly Arg Gly Arg Arg Leu Gly Gly Pro Arg
                   860                      865                      870
    Asp Pro Ala Pro Glu Lys Ser Arg Ala Ser Ser Ser Gly Gly Ser
                   875                      880                      885
    Gly Gly Gly Ser Gly Gly Arg Val Pro Lys Ser Ala Ser Val Ser
                   890                      895                      900
    Ala Leu Ser Leu Ile Ile Thr Ala Asp Asp Gly Ser Gly Gly Pro
                   905                      910                      915
    Leu Met Ser Pro Leu Ser Pro Arg Ser Leu Ser Ser Asn Pro Ser
                   920                      925                      930
    Ser Arg Asp Ser Ser Pro Ser Arg Asp Pro Ser Pro Val Cys Gly
                   935                      940                      945
    Ser Leu Arg Pro Pro Ile Val Ile His Ser Ser Gly Lys Lys Tyr
                   950                      955                      960
    Gly Phe Ser Leu Arg Ala Ile Arg Val Tyr Met Gly Asp Ser Asp
                   965                      970                      975
    Val Tyr Thr Val His His Val Val Trp Ser Val Glu Asp Gly Ser
                   980                      985                      990
```

9

PF-1688 P

```
Pro Ala Gln Glu Ala Gly Leu Arg Ala Gly Asp Leu Ile Thr His
        995               1000              1005
Ile Asn Gly Glu Ser Val Leu Gly Leu Val His Met Asp Val Val
        1010              1015              1020
Glu Leu Leu Leu Lys Ser Gly Asn Lys Ile Ser Leu Arg Thr Thr
        1025              1030              1035
Ala Leu Glu Asn Thr Ser Ile Lys Val Gly Pro Ala Arg Lys Asn
        1040              1045              1050
Val Ala Lys Gly Arg Met Ala Arg Arg Ser Lys Arg Ser Arg Arg
        1055              1060              1065
Arg Glu Thr Gln Asp Arg Arg Lys Ser Leu Phe Lys Lys Ile Ser
        1070              1075              1080
Lys Gln Thr Ser Val Leu His Thr Ser Arg Ser Phe Ser Ser Gly
        1085              1090              1095
Leu His His Ser Leu Ser Ser Ser Glu Ser Leu Pro Gly Ser Pro
        1100              1105              1110
Thr His Ser Leu Ser Pro Ser Pro Thr Thr Pro Cys Arg Ser Pro
        1115              1120              1125
Ala Pro Asp Val Pro Ala Asp Thr Thr Ala Ser Pro Pro Ser Ala
        1130              1135              1140
Ser Pro Ser Ser Ser Ser Pro Ala Ser Pro Ala Ala Ala Gly His
        1145              1150              1155
Thr Arg Pro Ser Ser Leu His Gly Leu Ala Ala Lys Leu Gly Pro
        1160              1165              1170
Pro Arg Pro Lys Thr Gly Arg Arg Lys Ser Thr Ser Ser Ile Pro
        1175              1180              1185
Pro Ser Pro Leu Ala Cys Pro Pro Ile Ser Ala Pro Pro Pro Arg
        1190              1195              1200
Ser Pro Ser Pro Leu Pro Gly His Pro Pro Ala Pro Ala Arg Ser
        1205              1210              1215
Pro Arg Leu Arg Arg Gly Gln Ser Ala Asp Lys Leu Gly Thr Gly
        1220              1225              1230
Glu Arg Leu Asp Gly Glu Ala Gly Arg Arg Thr Arg Gly Pro Glu
        1235              1240              1245
Ala Glu Leu Val Val Met Arg Arg Leu His Leu Ser Glu Arg Arg
        1250              1255              1260
Asp Ser Phe Lys Lys Gln Glu Ala Val Gln Glu Val Ser Phe Asp
        1265              1270              1275
Glu Pro Gln Glu Glu Ala Thr Gly Leu Pro Thr Ser Val Pro Gln
        1280              1285              1290
Ile Ala Val Glu Gly Glu Glu Ala Val Pro Val Ala Leu Gly Pro
        1295              1300              1305
Thr Gly Arg Asp
```

<210> 5
<211> 1331
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526158CD1

<400> 5
```
Met Lys Ser Arg Arg Asp Lys Leu His Ile Pro Ala Leu Thr Leu
 1               5                10                15
Asp Leu Ser Pro Ser Ser Gln Ser Pro Ser Leu Leu Gly Pro Ser
                20                25                30
```

PF-1688 P

```
Ser Pro Cys Ser Pro Cys Ser Pro Ser Leu Gly Leu His Pro Trp
            35                    40                    45
Ser Cys Arg Ser Gly Asn Arg Lys Ser Leu Val Val Gly Thr Pro
            50                    55                    60
Ser Pro Thr Leu Ser Arg Pro Leu Ser Pro Leu Ser Val Pro Thr
            65                    70                    75
Ala Gly Ser Ser Pro Leu Asp Ser Pro Arg Asn Phe Ser Ala Ala
            80                    85                    90
Ser Ala Leu Asn Phe Pro Phe Ala Arg Arg Ala Asp Gly Arg Arg
            95                   100                   105
Trp Ser Leu Ala Ser Leu Pro Ser Ser Gly Tyr Gly Thr Asn Thr
           110                   115                   120
Pro Ser Ser Thr Leu Ser Ser Ser Ser Ser Ser Arg Glu Arg Leu
           125                   130                   135
His Gln Leu Pro Phe Gln Pro Thr Pro Asp Glu Leu His Phe Leu
           140                   145                   150
Ser Lys His Phe Arg Ser Ser Glu Asn Val Leu Asp Glu Glu Gly
           155                   160                   165
Gly Arg Ser Pro Arg Leu Arg Pro Arg Ser Arg Ser Leu Ser Pro
           170                   175                   180
Gly Arg Ala Thr Gly Thr Phe Asp Asn Glu Ile Val Met Met Asn
           185                   190                   195
His Val Tyr Arg Glu Arg Phe Pro Lys Ala Thr Ala Gln Met Glu
           200                   205                   210
Gly Arg Leu Gln Glu Phe Leu Thr Ala Tyr Ala Pro Gly Ala Arg
           215                   220                   225
Leu Ala Leu Ala Asp Gly Val Leu Gly Phe Ile His His Gln Ile
           230                   235                   240
Val Glu Leu Ala Arg Asp Cys Leu Ala Lys Ser Gly Glu Asn Leu
           245                   250                   255
Val Thr Ser Arg Tyr Phe Leu Glu Met Gln Glu Lys Leu Glu Arg
           260                   265                   270
Leu Leu Gln Asp Ala His Glu Arg Ser Asp Ser Glu Glu Val Ser
           275                   280                   285
Phe Ile Val Gln Leu Val Arg Lys Leu Leu Ile Ile Ile Ser Arg
           290                   295                   300
Pro Ala Arg Leu Leu Glu Cys Leu Glu Phe Asp Pro Glu Glu Phe
           305                   310                   315
Tyr His Leu Leu Glu Ala Ala Glu Gly His Ala Arg Glu Gly Gln
           320                   325                   330
Gly Ile Lys Thr Asp Leu Pro Gln Tyr Ile Ile Gly Gln Leu Gly
           335                   340                   345
Leu Ala Lys Asp Pro Leu Glu Glu Met Val Pro Leu Ser His Leu
           350                   355                   360
Glu Glu Glu Gln Pro Pro Ala Pro Glu Ser Pro Glu Ser Arg Ala
           365                   370                   375
Leu Val Gly Gln Ser Arg Arg Lys Pro Cys Glu Ser Asp Phe Glu
           380                   385                   390
Thr Ile Lys Leu Ile Ser Asn Gly Ala Tyr Gly Ala Val Tyr Leu
           395                   400                   405
Val Arg His Arg Asp Thr Arg Gln Arg Phe Ala Ile Lys Lys Ile
           410                   415                   420
Asn Lys Gln Asn Leu Ile Leu Arg Asn Gln Ile Gln Gln Val Phe
           425                   430                   435
Val Glu Arg Asp Ile Leu Thr Phe Ala Glu Asn Pro Phe Val Val
           440                   445                   450
Ser Met Phe Cys Ser Phe Glu Thr Arg Arg His Leu Cys Met Val
           455                   460                   465
Met Glu Tyr Val Glu Gly Gly Asp Cys Ala Thr Leu Leu Lys Asn
```

11

PF-1688 P

```
                  470                 475                 480
    Met Gly Pro Leu Pro Val Asp Met Ala Arg Leu Tyr Phe Ala Glu
                  485                 490                 495
    Thr Val Leu Ala Leu Glu Tyr Leu His Asn Tyr Gly Ile Val His
                  500                 505                 510
    Arg Asp Leu Lys Pro Asp Asn Leu Leu Ile Thr Ser Leu Gly His
                  515                 520                 525
    Ile Lys Leu Thr Asp Phe Gly Leu Ser Lys Ile Gly Leu Met Ser
                  530                 535                 540
    Met Ala Thr Asn Leu Tyr Glu Gly His Ile Glu Lys Asp Ala Arg
                  545                 550                 555
    Glu Phe Ile Asp Lys Gln Val Cys Gly Thr Pro Glu Tyr Ile Ala
                  560                 565                 570
    Pro Glu Val Ile Phe Arg Gln Gly Tyr Gly Lys Pro Val Asp Trp
                  575                 580                 585
    Trp Ala Met Gly Val Val Leu Tyr Glu Phe Leu Val Gly Cys Val
                  590                 595                 600
    Pro Phe Phe Gly Asp Thr Pro Glu Glu Leu Phe Gly Gln Val Val
                  605                 610                 615
    Ser Asp Glu Ile Met Trp Pro Glu Gly Asp Glu Ala Leu Pro Ala
                  620                 625                 630
    Asp Ala Gln Asp Leu Ile Thr Arg Leu Leu Arg Gln Ser Pro Leu
                  635                 640                 645
    Asp Arg Leu Gly Thr Gly Gly Thr His Glu Val Lys Gln His Pro
                  650                 655                 660
    Phe Phe Leu Ala Leu Asp Trp Ala Gly Leu Leu Arg His Lys Ala
                  665                 670                 675
    Glu Phe Val Pro Gln Leu Glu Ala Glu Asp Asp Thr Ser Tyr Phe
                  680                 685                 690
    Asp Thr Arg Ser Glu Arg Tyr Arg His Leu Gly Ser Glu Asp Asp
                  695                 700                 705
    Glu Thr Asn Asp Glu Glu Ser Ser Thr Glu Ile Pro Gln Phe Ser
                  710                 715                 720
    Ser Cys Ser His Arg Phe Ser Lys Val Tyr Ser Ser Ser Glu Phe
                  725                 730                 735
    Leu Ala Val Gln Pro Thr Pro Thr Phe Ala Glu Arg Ser Phe Ser
                  740                 745                 750
    Glu Asp Arg Glu Glu Gly Trp Glu Arg Ser Glu Val Asp Tyr Gly
                  755                 760                 765
    Arg Arg Leu Ser Ala Asp Ile Arg Leu Arg Ser Trp Thr Ser Ser
                  770                 775                 780
    Gly Ser Ser Cys Gln Ser Ser Ser Ser Gln Pro Glu Arg Gly Pro
                  785                 790                 795
    Ser Pro Ser Leu Leu Asn Thr Ile Ser Leu Asp Thr Met Pro Lys
                  800                 805                 810
    Phe Ala Phe Ser Ser Glu Asp Glu Gly Val Gly Pro Gly Pro Ala
                  815                 820                 825
    Gly Pro Lys Arg Pro Val Phe Ile Leu Gly Glu Pro Asp Pro Pro
                  830                 835                 840
    Pro Ala Ala Thr Pro Val Met Pro Lys Pro Ser Ser Leu Ser Ala
                  845                 850                 855
    Asp Thr Ala Ala Leu Ser His Ala Arg Leu Arg Ser Asn Ser Ile
                  860                 865                 870
    Gly Ala Arg His Ser Thr Pro Arg Pro Leu Asp Ala Gly Arg Gly
                  875                 880                 885
    Arg Arg Leu Gly Gly Pro Arg Asp Pro Ala Pro Glu Lys Ser Arg
                  890                 895                 900
    Ala Ser Ser Ser Gly Gly Ser Gly Gly Gly Ser Gly Gly Arg Val
                  905                 910                 915
```

12

Copy provided by USPTO from the PACR

PF-1688 P

```
Pro Lys Ser Ala Ser Val Ser Ala Leu Ser Leu Ile Ile Thr Ala
        920             925             930
Asp Asp Gly Ser Gly Gly Pro Leu Met Ser Pro Leu Ser Pro Arg
        935             940             945
Ser Leu Ser Ser Asn Pro Ser Ser Arg Asp Ser Ser Pro Ser Arg
        950             955             960
Asp Pro Ser Pro Val Cys Gly Ser Leu Arg Pro Pro Ile Val Ile
        965             970             975
His Ser Ser Gly Lys Lys Tyr Gly Phe Ser Leu Arg Ala Ile Arg
        980             985             990
Val Tyr Met Gly Asp Ser Asp Val Tyr Thr Val His His Val Val
        995            1000            1005
Trp Ser Val Glu Asp Gly Ser Pro Ala Gln Glu Ala Gly Leu Arg
       1010            1015            1020
Ala Gly Asp Leu Ile Thr His Ile Asn Gly Glu Ser Val Leu Gly
       1025            1030            1035
Leu Val His Met Asp Val Val Glu Leu Leu Leu Lys Ser Gly Asn
       1040            1045            1050
Lys Ile Ser Leu Arg Thr Thr Ala Leu Glu Asn Thr Ser Ile Lys
       1055            1060            1065
Val Gly Pro Ala Arg Lys Asn Val Ala Lys Gly Arg Met Ala Arg
       1070            1075            1080
Arg Ser Lys Arg Ser Arg Arg Arg Glu Thr Gln Asp Arg Arg Lys
       1085            1090            1095
Ser Leu Phe Lys Lys Ile Ser Lys Gln Thr Ser Val Leu His Thr
       1100            1105            1110
Ser Arg Ser Phe Ser Ser Gly Leu His His Ser Leu Ser Ser Ser
       1115            1120            1125
Glu Ser Leu Pro Gly Ser Pro Thr His Ser Leu Ser Pro Ser Pro
       1130            1135            1140
Thr Thr Pro Cys Arg Ser Pro Ala Pro Asp Val Pro Ala Asp Thr
       1145            1150            1155
Thr Ala Ser Pro Pro Ser Ala Ser Pro Ser Ser Ser Ser Pro Ala
       1160            1165            1170
Ser Pro Ala Ala Ala Gly His Thr Arg Pro Ser Ser Leu His Gly
       1175            1180            1185
Leu Ala Ala Lys Leu Gly Pro Pro Arg Pro Lys Thr Gly Arg Arg
       1190            1195            1200
Lys Ser Thr Ser Ser Ile Pro Pro Ser Pro Leu Ala Cys Pro Pro
       1205            1210            1215
Ile Ser Ala Pro Pro Pro Arg Ser Pro Ser Pro Leu Pro Gly His
       1220            1225            1230
Pro Pro Ala Pro Ala Arg Ser Pro Arg Leu Arg Arg Gly Gln Ser
       1235            1240            1245
Ala Asp Lys Leu Gly Thr Gly Glu Arg Leu Asp Gly Glu Ala Gly
       1250            1255            1260
Arg Arg Thr Arg Gly Pro Glu Ala Glu Leu Val Val Met Arg Arg
       1265            1270            1275
Leu His Leu Ser Glu Arg Arg Asp Ser Phe Lys Lys Gln Glu Ala
       1280            1285            1290
Val Gln Glu Val Ser Phe Asp Glu Pro Gln Glu Glu Ala Thr Gly
       1295            1300            1305
Leu Pro Thr Ser Val Pro Gln Ile Ala Val Glu Gly Glu Glu Ala
       1310            1315            1320
Val Pro Val Ala Leu Gly Pro Thr Gly Arg Asp
       1325            1330
```

<210> 6
<211> 3912

PF-1688 P

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2509577CB1

<400> 6

```
gaagaattgt cggcattggc ctggccagcg ctggttggtg gcctggagga ggatttgatt 60
ggaaaaccaa cggtgcagct ggccgcggtg tccctgagaa aagtcagaaa tggccaatga 120
agcttttgct tataaaagga atgcgatgtt aattctgggg cattgatgtt ttacaatgcc 180
tgatcaagat aaaaaggtga agaccacaga aaaatcaact gataaacagc aagaaatcac 240
catcagggac tattcagatc ttaaaagact tcggtgcctt ttgaacgtcc aatcaagcaa 300
acaacagctt ccagccatta acttcgatag tgcccaaaat agcatgacga agtctgagcc 360
cgccatcagg gcgggtggac acagagctcg gggtcagtgg catgaatcca cagaagctgt 420
tgaacttgaa aattttagta taaactacaa gaatgagaga aatttcagca acatcctca 480
gcgtaaacta tttcaggaga tctttaccgc cttggtgaaa aatagactca taagcagaga 540
gtgggttaat cgagccccat ctattcattt tctgagagtg ttaatctgtc tgaggctact 600
aatgagggat ccatgttatc aggaaatact ccatagcttg ggtgggattg aaaacctagc 660
tcagtatatg gagattgtag ccaatgagta cctcggctat ggagaagagc agcacactgt 720
ggacaagctg gtcaacatga catatatttt tcaaaaactt gctgcagtca aagatcaaag 780
agaatgggtc accacaagtg gagcccacaa gacattagta aatttacttg gtgcccgaga 840
tactaatgtt ctattgggtt cccttctggc tctggctagt ttagcagaaa gtcaagaatg 900
tagggagaag ataagtgaac tcaacattgt agaaaatctg ttgatgattt tacatgaata 960
tgacttgctt tctaaaagac taacagcgga gttgctgcgc ctactttgtg cagagcccca 1020
ggtgaaagag caggtgaagc tctatgaggg gataccggtc ctcctcagtc tgctccactc 1080
tgaccacttg aagctcctct ggagcattgt ctggattctg gtacaggttt gtgaggaccc 1140
tgagaccagc gtggaaattc gcatttgggg aggcatcaaa cagcttcttc atattttaca 1200
aggagacaga aattttgttt ctgatcactc ctccattgga agcctgtcca gtgcaaatgc 1260
tgcaggccga atccagcagc ttcatttatc agaagacttg agccctaggg aaatacaaga 1320
aaatactttc tcacttcaag cagcctgctg tgctgccctc actgagctgg tgctcaatga 1380
caccaatgcc caccaggtgg ttcaggaaaa tggtgtatat acaatagcaa aattaatttt 1440
accaaataag caaaagaatg cagcaaaaag taatctatta cagtgttatg ctttcagagc 1500
cttgagattt ctcttcagta tggaagaaa cagaccactc tttaaaagac ttttccccac 1560
agacttgttt gagatcttca ttgacatagg gcattatgta cgtgatatca gtgcttatga 1620
agaattggta tccaagctga atttattagt ggaggatgaa ctgaagcaaa ttgctgaaaa 1680
tattgaaagc attaatcaga acaaagctcc tttgaaatat ataggcaact atgcaatttt 1740
ggatcatctt ggaagtggag cttttggctg tgtttacaag gttagaaagc atagtggtca 1800
aaatctttta gcaatgaaag aggtcaattt acataaccca gcatttggaa aggataagaa 1860
agatcgagac agcagcgtaa ggaatattgt ttctgaatta acaataatta aagagcagct 1920
ttatcatccc aacattgtac gttattacaa aacatttctg gaaaacgata ggttgtacat 1980
agttatggag ctgatagaag gagccccgct tggagagcat ttcagttctt tgaaggaaaa 2040
acatcaccat tttactgaag aaagactatg gaaaatattt atacagctgt gcttagctct 2100
tcgatactta cacaaggaga agaggattgt ccatagagat ctgacaccaa acaacattat 2160
gttggggggat aaggacaaag taacagttac tgactttggc ctggcaaagc aaaaacaaga 2220
aaacagtaaa ctcacgtctg tggttggaac aatcctgtat tcttccccg aggtactgaa 2280
gagtgagccg tatggggaga aggctgatgt ctgggcagta ggctgcatcc tttatcagat 2340
ggcgactttg agtccccct tctacagcac taacatgctg tccttggcta caaaaatagt 2400
ggaggcggta tatgaaccag tgccagaagg tatctactct gaaaaagtaa cagacaccat 2460
cagcaggtgc ctcactcctg atgcggaagc tcgtccagat attgtagaag tcagttcgat 2520
gatatcagat gtcatgatga aatatttaga caacttatct acatcccagt tgtccttgga 2580
aaagaagcta gaacgggaac gaagacgcac acaaaggtat tttatggaag ccaaccggaa 2640
caccgtcaca tgtcaccatg agctggctgt tctatctcac gagaccttttg agaaggcaag 2700
tttgagtagc agcagcagtg gagcagccag cctgaaaagt gaactttcag aaagcgcaga 2760
cctgccccct gaaggcttcc aggcctccta ggtaaagac gaagacaggg cctgtgacga 2820
aatcctgtca gatgataact tcaacctgga aaatgctgag aaagatacat attcagaggt 2880
agatgatgaa ttggacattt cggataactc cagcagctcc agttcaagcc ctctgaaaga 2940
atctacattc aacattttaa agagaagttt tagtgcttca ggaggagaaa gacaatccca 3000
aacaagggac ttcactggag gaacaggatc aagaccaaga ccagggccac agatgggcac 3060
```

14

```
attcttgtgg caagcatcag caggaattgc tgtgtcccag aggaaagtgc gtcagatcag 3120
tgatcctatt cagcagatat taattcagct gcacaaaata atctatatca cacagcttcc 3180
tccagctttg caccacaatt tgaaaagaag ggttatagag agattcaaga aatccctctt 3240
cagccagcag agtaaccctt gtaatttgaa atctgaaatt aaaaagttat ctcagggatc 3300
tccagaaccg attgagccca actttttcac agcagattac catttattac atcgttcatc 3360
cggtggaaac agcctgtccc caaatgaccc tacaggttta ccaaccagca ttgaattgga 3420
ggaaggaata acatatgaac agatgcagac tgtgattgaa gaagtccttg aggaaagtgg 3480
ctattacaat tttacatcta acaggtatca ttcctatcca tgggggacca agaatcaccc 3540
aaccaaaaga tgaaaatgct gcattttgag tggacttgat tttctcagtg aagttcaagt 3600
tctggacttc agccgctatt gcaagatgcc caaggattgg gtgctgctag agggtgtgga 3660
aaagaccaag atgccatggg gcctgcagga cttctttctg ggggtcctgt gctggagtat 3720
atgacagctg cggtacttga gggcttcatt gccagaacac attatataca ggatgtcaga 3780
gctaccagtg tgctgctggg agaaatgct gcaaaattca tcttttggag ggtggggga 3840
aaacccaaaa acaacaacaa aaaaactctc ttacagaatt ttccttaaca ttaaaaaaaa 3900
cttgtcatat tt                                                    3912
```

<210> 7
<211> 3229
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7505222CB1

<400> 7

```
gtagcataaa gaggaaatag aaggaaagca ataagtaaga aagtacatat ttcaatctga 60
aaatgcttgg cactactacc cttggaaaat gtagagaagt agccagtagc cgcgcctggg 120
gagtcgcctg aacgtgacgg cagcaaatgc agattgttgg gtctccggga ccaggagcag 180
cgtggccagt gaagcgcgtg gttttcccaa atggtgaaca attcttgtta tctgtggcca 240
caaagaaagt tatttgtctc tgtcttggca aggctgggag gaaagtttta gctaagaaac 300
tcagcccatt ggagaccatg gataagtacg atgtgattaa ggccatcggg caaggtgcct 360
tcgggaaagc atacttagct aaagggaaat cagatagcaa gcactgtgtc ataaaagaga 420
tcaattttga aaagatgccc atacaagaaa aagaagcttc aaagaaagaa gtgattcttc 480
tggaaaagat gaaacatccc aacattgtag ccttcttcaa ttcatttcaa gagaatggca 540
ggctgtttat tgtaatggaa tattgtgatg gaggggatct catgaaaagg atcaatagac 600
aacggggtgt gttatttagt gaagatcaga tcctcggttg gtttgtacag atttctctag 660
gactaaaaca tattcatgac aggaagatat tacacaggga cataaaagct cagaacattt 720
ttcttagcaa gaacggaatg gtggcaaagc ttggggactt tggtatagca agagtcctga 780
ataattccat ggaacttgct cgaacttgta ttggaacacc ttactacctg tccccagaga 840
tctgtcagaa taaaccctac aacaataaaa cggatatttg gtctcttggc tgtgtcttat 900
atgagctctg cacacttaaa catccttttg agggtaacaa cttacagcag ctggttctga 960
agatttgtca agcacatttt gccccaatat ctccggggtt ttctcgtgag ctccattcct 1020
tgatatctca gctctttcaa gtatctcctc gagaccgacc atccataaat tccattttga 1080
aaaggccctt tttagagaat cttattccca aatatttgac tcctgaggtc attcaggaag 1140
aattcagtca catgcttata tgcagagcag gagcgccagc ttctcgacat gctgggaagg 1200
tggtccagaa gtgtaaaata caaaaagtga gattccaggg aaagtgccca ccaagatcaa 1260
ggatatctgt gccaattaaa aggaatgcta tattgcatag aaatgaatgg agaccaccag 1320
ctggagccca gaaggccaga tctataaaaa tgatagaaag acccaaaatt gctgctgtct 1380
gtggacatta tgattattat tatgctcaac ttgatatgct gaggaggaga gcccacaaac 1440
caagttatca ccctattcct caagaaaata ctggagttga ggattacggt caggaaacga 1500
ggcatggtcc atccccaagt caatggcctg ctgagtacct tcagagaaaa tttgaagctc 1560
aacaatataa gttgaaagtg gagaagcaat gggtcttcg tccatcttct gccgagccaa 1620
attacaacca gagacaagag ctaagaagta atggagaaga gcctagattc caggagctgc 1680
catttaggaa aaacgaaatg aaggaacagg aatattggaa gcagttagag gaaatacgcc 1740
aacagtacca caatgacatg aaagaaatta gaaagaagat g gggagagaa ccagaggaga 1800
actcaaaaat aagtcataaa acctatttgg tgaagaagag taacctgcct gtccatcaag 1860
atgcatctga gggagaagca cctgtgcagg acattgaaaa agacttgaaa caaatgaggc 1920
ttcagaacac aaaggaaagt aaaaatccag aacagaaata taaagctaag ggggtaaaat 1980
```

PF-1688 P

```
ttgaaattaa tttagacaaa tgtatttctg atgaaaacat cctccaagag gaagaggcaa 2040
tggatatacc aaatgaaact ttgacctttg aggatggcat gaagtttaag gaatatgaat 2100
gtgtaaagga gcatggagat tatacagaca aagcatttga aaaacttcac tgcccagaag 2160
cagggttttc cacgcagact gtagctgctg tgggaaacag gaggcagtgg gatggaggag 2220
cgcctcagac tctgctgcag atgatggcag tggccgacat cacctccacc tgccccacgg 2280
ggcctgacaa tggccaagtt attgtgattg aaggcattcc aggaaacagg aaacagtggc 2340
ggcatgaagc tccaggaact ttaatgagtg ttttggcagc agcacatcta acgagtagct 2400
cattttctgc cgatgaagaa tttgcaatgg gaacattaaa acaatggcta cccaaagaag 2460
aagatgaagg gaaggtagaa atggtctctg gcattgaagt agatgaggaa caactagaac 2520
caagatctga tgatgatgat acaaattttg aagaatctga agatgagttg agagatgaag 2580
tagtagaata cttagaaaaa ctcgctactt tcaaagggga agaaaaaaca gaagaggcct 2640
ccagtacctc taaggactct agaaagtcaa gagaaagaga ggggataagt atgcagaaat 2700
ctgaagaatt aagggagggc ttggagaata tttctactac atctaatgac cacatttgta 2760
ttactgatga agaccaagga acatcaacaa ccagtcaaaa tatacaagtg tgattattgt 2820
actttttctt aagtaataag ttagtgtcta ttacctatag tatttatttg ggtacaagtc 2880
ataaatgctc atttactgta agggttttct agtaatctca aggatttatt aatttttctt 2940
tcaatttagg aagtagaact tttgaatata gccattaata tttttacttt aaagtttcta 3000
ttaagaaatc ttaggccggg cagtctcatc actttgggag gccaaggcag gcagatcatg 3060
aggtcaggag ttgagaccag tccaaccaac atggtgaaac ccgtctcta ctaaaaatac 3120
aaaattagct gggcatggtg gtgcatgcct gtaatcccag ttacttggga ggctgaggca 3180
ggagaatcac ctgaacccag gagatggaag ttgcaagtga gccgagatg·           3229
```

<210> 8
<211> 2100
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524408CB1

<400> 8

```
tagagcattg ccttgttgct gacctttcag tatggggagg attggcatct cctgtctttt 60
tcctgcttct tggcatttta gcatatctcc agtagggtgt cctcgaattc tgaataccaa 120
tttacgccaa attatggtca ttagtgtcct ggctgctgct gtttcacttt tatattttc 180
tgttgtcata atccgaaata agtatgggcg actaaccaga gacaagaaat tcaaaggta 240
cctggcacga gttaccgaca ttgaagctac agacaccaat aaccccaatg tgaactatgg 300
gatcgtggtg gactgtggta gcagtgggtc tcgagtattt gtttactgct ggccaaggca 360
taatggcaat ccacatgatc tgttggatat caggcaaatg agggataaaa accgaaagcc 420
agtggtcatg aagataaaac cgggcatttc agaatttgct acctctccag agaaagtcag 480
tgattacatt tctccacttt tgaactttgc tgcagagcat gtgccacggg caaaacacaa 540
agagacacct ctctacattc tctgcacggc tggaatgaga atcctcccg aaagccagca 600
gaaagctatt ctggaagacc ttctgaccga tatccccgtg cactttgact ttctgttttc 660
tgactctcat gcagaagtaa tttctgggaa acaagaaggt gtgtatgctt ggattggcat 720
taatttttgtc cttggacgat ttgagcatat tgaagatgat gatgaggccg ttgtggaagt 780
taacattcct ggaagtgtaa gcagcgaagc cattgtccgt aaaaggacag cgggcattct 840
cgacatgggc ggcgtgttga ctcagatagc gtacgaagtc cccaaaactg caagctttgc 900
gtcctcacag caggaagaag tagctaaaaa cttgttagct gaatttaact gggatgtga 960
tgttcaccaa actgagcatg tgtatcgagt ctatgtggcc acgttttttg ggtttggtgg 1020
caatgctgct cgacagagat aatatttgcc aacaccattc aaaagaacag 1080
gctcctgggt aaacagactg gtctgactcc tgatatgccg tacttggacc cctgcctacc 1140
cctagacatt aaagatgaaa tccagcaaaa tggacaaacc atataacctac gagggactgg 1200
agactttgac ctgtgtcgag agactatcca gcctttcatg aataaaacaa acgagaccca 1260
gacttccctc aatggggtct accagccccc aattcacttc cagaacagtg aattctatgg 1320
cttctccgaa ttctactact gcaccgagga tgtgttacga atggggggag actacaatgc 1380
tgctaaattt actaaagctg caaaggatta ttgtgcaaca aagtggtcca ttttgaactgc 1440
acgctttgac cgaggactgt acgcctctca tgctgacctc cacaggctta agtgaactgc 1500
tccatgttgt gggaccaggg atgtgaagca agtacatcaa ccttgaaacg cgcgcagttg 1560
gtttgggaga gccggcctga ggaagcccct gccccaaggc tgcccacaga gggaaggatt 1620
```

16

PF-1688 P

```
gtgtgtgtgt gtgtgtgtgt gccatcttga cagagtgagt cagtctgact tgcctttgtg 1680
cttgccgttt gtaggtatca gtgcttcaaa tcggcctgga tgtttgaggt gtttcatagg 1740
ggcttttcgt ttcctgtcaa ctataaaagc ttaaagactg ccttgcaagt ttacgacaag 1800
gaggttcagt ggacccttgg agccatcctc tacaggaccc gctttctacc attaagagac 1860
atccagcagg aggccttccg agccagtcac acccactggc ggggcgtttc ctttgtctac 1920
aaccactacc tgttctctgg ctgcttcctg gtggtgctgc tggccatcct gctgtacctg 1980
ctgcggctgc ggcgcatcca caggcgcact ccccggagca gctcggccgc cgccctctgg 2040
atggaggagg gccttcccgc ccagaatgcc ccggggacct tgtgatccag ctcacagcta 2100
```

<210> 9
<211> 4213
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526163CB1

<400> 9
```
cccgggccat ggacgagtcg agcctcctgc ggcgccgcgg gctccagaag gagctgagcc 60
tgccacgccg aggacgtggc tgccgcagcg ggaaccgcaa gagcttggtg gtaggaacgc 120
cctccccgac cctctcccgg cccctgtcgc cattgtcggt cccaacggca ggcagcagcc 180
ccttggatag tcctcggaat ttctcggctg cctctgccct aaatttcccc tttgcccgga 240
gggcagacgg cagaagatgg tccctcgcgt ctctcccatc ttccggctat ggaaccaaca 300
cacccagctc caccctctcg tcaagctcat cctcccggga acgtctccac cagcttccct 360
tccagccgac gccggacgag ctgcacttcc tgtccaagca cttccgcagc tcagagaatg 420
tgcttgatga ggaaggcggc cggtcacccc gcctccgacc ccgctctcgc agtctcagcc 480
cgggccgtgc aacggggacc ttcgacaatg agattgtcat gatgaatcac gtgtaccggg 540
agaggttccc caaggccaca gcacagatgg agggccgtct gcaggagttc ctgacggcct 600
acgcgcccgg cgcccggctg gcgctggctg atggcgtctt gggcttcatc caccaccaga 660
tcgtcgagct ggcccgagac tgcttggcca agtctggcga gaacctcgtc acctcccgct 720
acttcctaga gatgcaggag aagctggagc ggcttctgca ggatgcccat gagcgttcgg 780
acagtgagga ggtcagcttc atcgtccagc ttgtccggaa actgctgatc atcatctcac 840
ggccagctcg gctgctggag tgtctggagt ttgaccctga ggaattttac cacctgctgg 900
aggcggctga gggccatgcg cgggagggcc aaggcattaa gactgacctt ccacagtaca 960
tcattgggca gctgggcctg gccaaggacc ccctggagga gatggtgcca ctgagtcacc 1020
tcgaagaaga acagccccca gcacctgagt ccccagagag ccgcgccctg gtcggccagt 1080
cacggaggaa gccatgcgaa agcgactttg agaccatcaa actcattagc aacggagcct 1140
atggggccgt ctacctggtg cggcaccgtg acacacggca gcgctttgcc atcaagaaga 1200
tcaacaaaca gaacttgatc ctgcgtaacc aggtccagca ggtctttgtg gagcgtgaca 1260
ttctcacctt tgccgagaac ccctttgtgg tcagcatgtt ctgctccttt gagacccggc 1320
gccacctatg tatggtcatg gaatacgtgg aaggcggcga ctgcgccacg ctcctgaaga 1380
acatgggccc gctgcccgtg gacatggccc gcctgtactt cgccgagacg gtgttggcgc 1440
tggagtacct gcataactat ggcatcgtgc accgtgacct caaaccagac aatctgctca 1500
tcacctcgct tggccacatc aagctcacgg acttcggcct gtccaagatc ggcctcatga 1560
gcatggccac caacctctat gagggccaca tcgagaagga cgcccgagag ttcatcgaca 1620
agcaggtgtg tgggacgccg gagtacatag cccccgaggt gatcttccgc cagggctatg 1680
ggaagccagt ggactggtgg gccatgggcg tcgtcctcta tgagtttctg gtgggctgcg 1740
tgcctttctt tggagatacc cccgaggaac tcttcggtca ggtggtcagc gatgagatca 1800
tgtggccaga gggagatgag gcccttccag cagacgccca ggacctcatc accaggttgc 1860
tccggcagag cccgctggac cgtctgggca ctggtggcac cacgaagtg aagcagcacc 1920
cctttttcct ggccctggac tgggcagggc ttctccgaca caaagccgag ttcgtgcccc 1980
agctcgaagc cgaggatgat accagctact ttgacacacg ttcggaacgt taccgccatc 2040
tgggctccga ggacgacgag accaatgatg aagaatcgtc cacagagatc ccccagttct 2100
cctcctgctc ccaccggttc agcaaggtct acagcagctc tgagttcctg gccgtccagc 2160
ccactcctac cttcgctgaa aggagcttca gtgaagaccg ggaggagggg tgggagcgca 2220
gcgaagtgga ctatggccgc cggctgagtg ctgacatccg gctgaggtcc tggacatcct 2280
ctggatcctc ctgtcagtca tcttcgtccc agcccgagcg gggtcccagc ccatctctcc 2340
tgaataccat cagcctggac acaatgccca gtttgccttt ctcatcagag gatgaggggg 2400
```

17

PF-1688 P

```
taggcccagg ccctgcaggc cccaagaggc ccgtcttcat tctaggggag cctgacccc 2460
caccagcggc caccccagtg atgcccaagc cctcgagcct ttctgccgac acagctgctc 2520
tcagccacgc ccgcctacgg agcaatagca tcggcgcccg acactccaca ccaaggcctc 2580
tggatgccgg ccggggccgc cgccttgggg gcccaagaga cccagcccct gagaagtcca 2640
gagcctcctc cagcggtggc agtggtggcg gcagtggggg ccgcgtgccc aagtcagcct 2700
ctgtctctgc cctgtccctc atcatcacgg cagatgatgg cagcggcggc cccctcatga 2760
gcccccttc cccgcgctct ctgtcctcga acccgtcgtc ccgtgactct tcgccgagcc 2820
gagacccgtc ccccgtgtgt ggcagcctgc ggcccccat cgttatccac agctctggca 2880
agaagtacgg cttcagcctg cgggcgatcc gcgtctacat gggtgatagc gacgtctaca 2940
ctgtgcacca cgtcgtctgg agtgtggagg acggaagccc cgcccaggag gcgggcctgc 3000
gggctgggga cctcatcacc cacatcaacg gggagtcagt gctggggctg gtgcacatgg 3060
acgtcgtgga gctgctgctg aagagcggca acaagatatc cctgcggacc acagccctgg 3120
agaacacctc catcaaggtg ggccccgccc ggaagaatgt ggccaagggc cgcatggcac 3180
gcaggagcaa gaggagccgt cggcgggaga cccaggatcg gcggaagtca cttttcaaga 3240
agatctccaa gcagacctcc gtgctgcaca ccagccgcag cttctcctcc ggactccacc 3300
actcactgtc atccagtgag agcctccccg gctcgcccac ccacagcctc tcccccagcc 3360
ccaccactcc ctgccgaagc ccagcccctg atgtcccagc agataccact gcatccccac 3420
ccagcgcatc cccgagctcc agcagccccg cctcccagc tgctgctggc cacacccgcc 3480
ccagctccct gcacggcctg gctgccaagc ttgggccacc ccgccccaag actggccgcc 3540
gcaagtccac cagcagcatc ccgccctccc cgctggcctg cccgcccatc tccgcgcccc 3600
cacccccgctc gccctcgccc ctgcccgggc acccgcccgc acctgcccga tccccgcggc 3660
tgcgccgggg ccagtcagct gacaagctgg gcacagggga gcggctggat ggggaggcgg 3720
ggcggcgcac tcgtgggcca gaggccgagc tcgtggtcat gcggcggctg cacctgtccg 3780
agcgccgaga ctccttcaag aagcaggagg ccgtgcagga ggttagcttc gatgagccgc 3840
aggaggaggc cactgggctg cccacctcag tgccacagat cgccgtggag ggcgaggaag 3900
ccgtgccagt agctctcggg cccaccggaa gagactgatc ccctgccagg tctctccctg 3960
gcatcaaagt tacgcgtttt cttgtgcaat gtttttccg taaagtcatg cctggatggg 4020
gactgagcca ccagcctgac acccagaagg cgagaagcca tctcggtcct tgctggaagg 4080
tggagacatc gcttgtgttc tggtgtcaat caggggctg gatggggcaa gaatggggga 4140
caagggtggc tttgtaaata gcagcaaatc cctgcaacta atttattact ttttctttt 4200
tttttttttt ttt                                                  4213
```

<210> 10
<211> 5991
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526158CB1

<400> 10

```
gtaccccgcc gcacagaggc ggcctctgcc tcggcatgaa gtcccgcagg gacaagctgc 60
acatcccggc gctgaccctc gatctgtctc cgagcagcca gagcccgtcc ctgctgggtc 120
ccagcagccc ctgcagcccc tgtagccccct ccttgggcct gcaccctgg agctgccgca 180
gcgggaaccg caagagcttg gtggtaggaa cgccctcccc gaccctctcc cggcccctgt 240
cgccattgtc ggtcccaacg gcaggcagca gcccccttgga tagtcctcgg aatttctcgg 300
ctgcctctgc cctaaatttc ccctttgccc ggagggcaga cggcagaaga tggtccctcg 360
cgtctctccc atcttccggc tatggaacca acacacccag ctccaccctc tcgtcaagct 420
catcctctcg ggaacgtctc caccagcttc ccttccagcc gacgccggac gagctgcact 480
tcctgtccaa gcacttccgc agctcagaga atgtgcttga tgaggaaggc ggccggtcac 540
cccgcctccg accccgctct cgcagtctca gcccgggccg tgcaacgggg accttcgaca 600
atgagattgt catgatgaat catgtgtacc gggagaggtt ccccaaggcc acagcacaga 660
tggagggccg tctgcaggag ttcctgacgg cctacgcgcc cggcgcccgg ctggcgctgg 720
ctgatggcgt cttgggcttc atccaccacc agatcgtcga gctggcccga gactgcttgg 780
ccaagtctgg cgagaacctc gtcacctccc gctacttcct agagatgcag gagaagctgg 840
agcggcttct gcaggatgcc catgagcgtt cggacagtga ggaggtcagc ttcatcgtcc 900
agcttgtccg gaaactgctg atcatcatct cacggccagc tcggctgctg gagtgtctgg 960
agtttgaccc tgaggaattt taccacctgc tggaggcggc tgagggccat gcgcggggagg 1020
```

```
gccaaggcat taagactgac cttccacagt acatcattgg gcagctgggc ctggccaagg 1080
accccctgga ggagatggtg ccactgagtc acctcgaaga agaacagccc ccagcacctg 1140
agtccccaga gagccgcgcc ctggtcggcc agtcacggag gaagccatgc gaaagcgact 1200
ttgagaccat caaactcatt agcaacggag cctatggggc cgtctacctg gtgcggcacc 1260
gtgacacacg gcagcgcttt gccatcaaga agatcaacaa acagaacttg atcctgcgta 1320
accagatcca gcaggtcttt gtggagcgtg acattctcac ctttgccgag aacccctttg 1380
tggtcagcat gttctgctcc tttgagaccc ggcgccacct atgtatggtc atggaatacg 1440
tggaaggcgg cgactgcgcc acgctcctga gaacatgggg cccgctgccc gtggacatgg 1500
cccgcctgta cttcgccgag acggtgttgg cgctggagta cctgcataac tatggcatcg 1560
tgcaccgtga cctcaaacca gacaatctgc tcatcacctc gcttggccac atcaagctca 1620
cggacttcgg cctgtccaag atcggcctca tgagcatggc caccaacctc tatgagggcc 1680
acatcgagaa ggacgcccga gagttcatcg acaagcaggt gtgtgggacg ccggagtaca 1740
tagccccga ggtgatcttc cgccagggct atgggaagcc agtggactgg tgggccatgg 1800
gcgtcgtcct ctatgagttt ctggtgggct gcgtgccttt ctttggagat accccgagg 1860
aactcttcgg tcaggtggtc agcgatgaga tcatgtggcc agagggagat gaggcccttc 1920
cagcagacgc ccaggacctc atcaccaggt tgctccggca gagcccgctg accgtctgg 1980
gcactggtgg cacccacgaa gtgaagcagc acccctttt cctggccctg gactgggcag 2040
ggcttctccg acacaaagcc gagttcgtgc cccagctcga agccgaggat gataccagct 2100
actttgacac acgttcggaa cgttaccgcc atctgggctc cgaggacgac gagaccaatg 2160
atgaagaatc gtccacagag atcccccagt tctcctcctg ctcccaccgg ttcagcaagg 2220
tctacagcag ctctgagttc ctggccgtcc agcccactcc taccttcgct gaaaggagct 2280
tcagtgaaga ccgggaggag gggtgggagc gcagcgaagt ggactatggc cgccggctga 2340
gtgctgacat ccggctgagg tcctggacat cctctggatc ctcctgtcag tcatcttcgt 2400
cccagcccga gcggggtccc agcccatctc tcctgaatac catcagcctg gacacaatgc 2460
ccaagtttgc cttctcatca gaggatgagg gggtaggccc aggccctgca ggccccaaga 2520
ggcccgtctt cattctaggg gagcctgacc ccccaccagc ggccacccca gtgatgccca 2580
agccctcgag cctttctgcc gacacagctg ctctcagcca cgcccgccta cggagcaata 2640
gcatcggcgc ccgacactcc acaccaaggc ctctggatgc cggccggggc cgccgccttg 2700
ggggcccaag agacccagcc cctgagaagt ccagagcctc ctccagcggt ggcagtggtg 2760
gcggcagtgg gggccgcgtg cccaagtcag cctctgtctc tgccctgtcc ctcatcatca 2820
cggcagatga tggcagcggc ggccccctca tgagcccct ttccccgcgc tctctgtcct 2880
cgaacccgtc gtcccgtgac tcttcgccga gccgagaccc gtccccgtg tgtggcagcc 2940
tgcggccccc catcgttatc cacagctctg gcaagaagta cggcttcagc ctgcgggcga 3000
tccgcgtcta catgggtgat agcgacgtct acactgtgca ccacgtcgtc tggagtgtgg 3060
aggacggaag ccccgcccag gaggcgggcc tgcgggctgg ggacctcatc acccacatca 3120
acggggagtc agtgctgggg ctggtgcaca tggacgtcgt ggagctgctg ctgaagagcg 3180
gcaacaagat atccctgcgg accacagccc tggagaacac ctccatcaag gtgggccccg 3240
cccggaagaa tgtggccaag ggccgcatgg cacgcaggag caagaggagc cgtcggcggg 3300
agacccagga tcggcggaag tcactttca agaaagatctc caagcagacc tccgtgctgc 3360
acaccagccg cagcttctcc tccggactcc accactcact gtcatccagt gagagcctcc 3420
ccggctcgcc cacccacagc ctctccccca gccccaccac tccctgccga agcccagccc 3480
ctgatgtccc agcagatacc actgcatccc cacccagcgc atccccgagc tccagcagcc 3540
ccgcctcccc agctgctgct ggccacaccc gccccagctc cctgcacggc ctggctgcca 3600
agcttgggcc accccgcccc aagactggcc gccgcaagtc caccagcagc atcccgccct 3660
cccgctggc ctgcccgccc atctccgcgc ccccacccg ctcgccctcg ccctgcccg 3720
ggcacccgcc cgcacctgcc cgatccccgc ggctgcgccg gggccagtca gctgacaagc 3780
tgggcacagg ggagcggctg gatggggagg cggggcggcg cactcgtggg ccagaggccg 3840
agctcgtggt catgcggcgg ctgcacctgt ccgagcgccg agactccttc aagaagcagg 3900
aggccgtgca ggaggttagc ttcgatgagc cgcaggagga ggccactggg ctgcccacct 3960
cagtgccaca gatcgccgtg gagggcgagg aagccgtgcc agtagctctc gggcccaccg 4020
gaagagactg atccctgcc aggtctctcc ctggcatcaa agttacgcgt tttcttgtgc 4080
aatgtttttt ccgtaaagtc atgcctggat ggggactgag ccaccagcct gacacccaga 4140
aggcgagaag ccatctcggt ccttgctgga aggtggagac atcgcttgtg ttctggtgtc 4200
aatcaggggg ctggatgggg caagaatggg ggacaagggt ggctttgtaa atagcagcaa 4260
atccctgcaa ctaatttatt actttttttt tttttttttt tttttttttt tgagacagag 4320
tctcactctg ttgccgggc tggagtgcag cggcgtgatc tcagctcact gcaacctccg 4380
cctcccaagt tcaagcgatt gtcctgcctc agcttcccaa gtggctggga ttacaggcgc 4440
ccaccactat gcccagctaa tttttgtat ttttagtaca gacggggttt caccatgttg 4500
gtcaggctgg tctcgaactc ctgacctcat gatttgcctg cctttgcctc ccaaagtgct 4560
```

19

PF-1688 P

```
gggattacag gcgtgagcca ctgggcccag cctaatttat tacttttat aagcgatagc 4620
cgtactgagc cgcccctga aggcggctgc caggtcttgc cccaggcacc tgggactctg 4680
tttgcaggcc ctgccctctg ggctgagaag gatgcacttt ggacaagtca tctgtgtttg 4740
tgttttccag ttttctgta cttttaagt gttttgtgtt acctggtctc attccctcc 4800
ccacacctac ccatttgagg ggatggagtt gaagtcacct ggtcacctgt accggcccag 4860
ttcggctaca acctggagtg tccgtaaaca attcctctca cccacaaaac aatgtaatcc 4920
cagcgatgga ctggattctg aaggccactt cccaccatca tagctgccat gcccaggcag 4980
tgcctgctct atatatagag tctgcctcca atcctgctgg cttcagcctg gagaagggat 5040
atgggagctg gagctttgat ggatgaatag gtgttcaccg gatctgggca gaggggtcat 5100
ccgctcccca ggtgggcact gataaaggaa ggtacaggcc tcacctggaa ctgccaaggc 5160
agcctccaga aatgctcggc tgtctcgggg cacgctccag tatgccagtc ctgcgggatt 5220
acgtccagct acttccagaa acactcagtg tcccctcccc tcaggctctg ccttggcctg 5280
gccttgtcca gtctaccctg gacaagatgc cgtgtgtttg aggcccagca gagtaagccc 5340
ttggccgtga tgtgtctgaa acacctgtta ggggttccct ccatatgtca gagcctctct 5400
gggatgaagt tcaagccaga aaacccagtc gaggctcaag tttgaatttc agcttcactg 5460
tgtggctctg ggaaaatggc tttcccactc tgtgcctcag tttccttgtg tttacaagac 5520
taatcccatt gactgtttat taagcaccta ctgtgtgcca agcgctttta cgtggcttct 5580
ccctcagcca gccttgagaa ggctggaggt ggtgtcatca cctccatttt acagacaaag 5640
cagctgagac cccagcgagg ggcggagacc tgtcccacga tcacccagca ggagtcgtgg 5700
cagaacggag catcagccag accctgttgt gggcgttgtc atcaagggag cttgaatgga 5760
gggtctggtg tcagatacag ccgactccag ccccagctca tcccccatga tgctgtgtga 5820
cccactgggc actctggtga gggagctttc cagacatcaa cagcccactc tgcttccctt 5880
tctgagtccc ctgtccagca ctgcctagtg ttggagggta gaccaaggct gtgcatgatt 5940
cacccctcc ttccatcctg gagctggcag tgaataaaag cccgtattta c         5991
```

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2509577	1	2509577CD1	6	2509577CB1
7505222	2	7505222CD1	7	7505222CB1
7524408	3	7524408CD1	8	7524408CB1
7526163	4	7526163CD1	9	7526163CB1
7526158	5	7526158CD1	10	7526158CB1

PF-1688 P

1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	2509577CD1	g10312094	6.4E-40	NIMA-related serine/threonine kinase [Mus musculus]
				Kandli, M. et al. Isolation and characterization of two evolutionarily conserved murine kinases (Nek6 and nek7) related to the fungal mitotic regulator, NIMA. Genomics 68, 187-196 (2000)
1	2509577CD1	714317|p1044.3	2.2E-130	[Caenorhabditis elegans] Protein containing 16 EB module domains and a protein kinase domain, has a region of low similarity to NIMA (never in mitosis gene a) - related expressed kinase 6 (human NEK6), which activates the S6 ribosomal protein kinase p70S6K (RPS6KB1)
				Chervitz, S. A. et al. Comparison of the complete protein sets of worm and yeast: Orthology and divergence. Science 282, 2022-2028 (1998)
1	2509577CD1	78975l|NEK6	1.3E-40	[Homo sapiens][Protein kinase;Transferase] NIMA (never in mitosis gene a) - related expressed kinase 6, a protein kinase that phosphorylates and activates the S6 ribosomal protein kinase p70S6K (RPS6KB1)
				Belham, C. et al. Identification of the NIMA family kinases NEK6/7 as regulators of the p70 ribosomal S6 kinase. Curr. Biol. 11, 1155-1167 (2001)
2	7505222CD1	g256855	5.4E-112	serine/threonine- and tyrosine-specific protein kinase; Nek1 [Mus sp.]
				Letwin, K. et al. A mammalian dual specificity protein kinase, Nek1, is related to the NIMA cell cycle regulator and highly expressed in meiotic germ cells. EMBO J. 11, 3521-3531 (1992)

PF-1688 P

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
2	7505222CD1	750718|NEK1	2.3E-110	[Homo sapiens] Protein containing a protein kinase domain, has weak similarity to serine threonine kinase 2 (mouse Stk2), which undergoes cleavage by caspase 3 (mouse Casp3) and the released N-terminal kinase domain and C-terminal domain promote apoptosis
				Nagase, T. et al. Prediction of the coding sequences of unidentified human genes. XXI. The complete sequences of 60 new cDNA clones from brain which code for large proteins. DNA Res. 8, 179-187 (2001)
2	7505222CD1	430066|Nek3	2.0E-92	[Mus musculus][Protein kinase;Transferase][Cytoplasmic] NIMA-related kinase 3, a protein kinase that is involved in cell cycle control
				Chen, A. et al. NIMA-related kinases: isolation and characterization of murine nek3 and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3. Gene 234, 127-137 (1999)
3	7524408CD1	g4583675	3.7E-264	[Homo sapiens] apyrase
				Biederbick, A. et al. A human intracellular apyrase-like protein, LALP70, localizes to lysosomal/autophagic vacuoles. J. Cell. Sci. 112 (Pt 15), 2473-2484 (1999)
3	7524408CD1	340820|LYSAL1	2.7E-265	[Homo sapiens][Other phosphatase;Hydrolase][Lysosome/vacuole;Cytoplasmic] Lysosomal apyrase-like protein (Golgi apyrase), a member of the apyrase or GDA1/CD39 family that is a lysosomal membrane protein with four apyrase domains, alternative splice form is identical to uridine diphosphatase

2

PF-1688 P

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Biederbick, A. et al. First apyrase splice variants have different enzymatic properties. J. Biol. Chem. 275, 19018-19024 (2000)
3	7524408CD1	753913|Lysal2	2.8E-160	[Mus musculus] Protein with high similarity to lysosomal apyrase-like protein (Golgi apyrase, human LYSAL1), which is a lysosomal membrane protein with four apyrase domains, member of the GDA1 or CD39 family of nucleoside phosphatases
				Shi, J. D. et al. Molecular cloning and characterization of a novel mammalian endo-apyrase (LALP1). J. Biol. Chem. 276, 17474-17478 (2001)
4	7526163CD1	g13537204	0.0	MAST205 [Homo sapiens]
4	7526163CD1	423529|KIAA0561	0.0	[Homo sapiens][Protein kinase;Transferase] Protein with high similarity to murine Mtssk, which is a protein kinase that interacts with microtubules and facilitates their organization in spermatids, contains a eukaryotic protein kinase domain and a PDZ domain
				Nagase, T. Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res. 5, 31-39 (1998)
4	7526163CD1	742582|MAST205	0.0	[Homo sapiens][Protein kinase;Transferase][Cytoskeletal] Protein with strong similarity to microtubule associated testis specific serine/threonine protein kinase (mouse Mtssk), which may act in spermatid maturation and microtubule organization, contains a protein kinase domain and a PDZ, DHR, or GLGF domain
				Walden, P. D. et al. Increased activity associated with the MAST205 protein kinase complex during mammalian spermiogenesis. Biol. Reprod. 55, 1039-1044 (1996)

3

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
5	7526158CD1	g406058	0.0	protein kinase [Mus musculus] Walden, P. D. et al. A novel 205-kilodalton testis-specific serine/threonine protein kinase associated with microtubules of the spermatid manchette. Mol. Cell. Biol. 13, 7625-7635 (1993)
5	7526158CD1	423529|KIAA0561	0.0	[Homo sapiens][Protein kinase;Transferase] Protein with high similarity to murine Mtssk, which is a protein kinase that interacts with microtubules and facilitates their organization in spermatids, contains a eukaryotic protein kinase domain and a PDZ domain; Nagase, T. DNA Res. 5, 31-39 (1998) supra
5	7526158CD1	582149|Mtssk	0.0E+00	[Mus musculus][Protein kinase;Transferase][Cytoplasmic;Cytoskeletal] Microtubule associated testis specific serine/threonine protein kinase, may be involved in the organization of manchette microtubules in spermatids, may have a role in spermatid maturation; Lumeng, C. et al. Interactions between beta 2-syntrophin and a family of microtubule-associated serine/threonine kinases. Nat. Neurosci. 2, 611-617 (1999)

4

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2509577CD1	1125	Armadillo/beta-catenin-like repeat; V198-E238, S239-A279, E280-E320, N401-S448	HMMER_PFAM
			Protein kinase domain: Y519-I783	HMMER_PFAM
			Armadillo/beta-catenin-like repeats:A197-E238, C278-E320, N401-S448	HMMER_SMART
			Serine/Threonine protein kinases, catalyti: Y519-L791	HMMER_SMART
			Tyrosine kinase, catalytic domain: Y519-I783	HMMER_SMART
			Receptor tyrosine kinase class II IPB002011: I578-K622, I651-K702, L735-V779	BLIMPS_BLOCKS
			Protein kinases signatures and profile: R630-S684	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: L645-L663, A710-T732, Y754-I776	BLIMPS_PRINTS
			SERINE/THREONINE PROTEIN KINASE D1044.3 IN CHROMOSOME III EC 2.7.1. PROTEIN TRANSFERASE ATP-BINDING EGF-LIKE DOMAIN PD140750:E108-I334 D775-Q813 C319-G517	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004 |P11837|13-285:I521-H646 I651-D775 |P41951|433-687: I521-I776 |P51954|6-248: L522-I776 |P51955|10-261: I521-I776	BLAST_DOMO

1

PF-1688 P

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S40, S235, S266, S376, S479, S540, S567, S618, S696, S755, S852, S900, S935, S974, S1039, S1068, T11, T15, T22, T58, T270, T324, T593, T670, T812, T898, T1077, T1092, T1123, Y746	MOTIFS
			Potential Glycosylation Sites: N86, N96, N187, N401, N793, N911, N1105	MOTIFS
			Protein kinases ATP-binding region signature: L525-K548	MOTIFS
			Tyrosine protein kinases specific active-site signature: I651-L663	MOTIFS
2	7505222CD1	888	Protein kinase domain: Y61-L316	HMMER_PFAM
			Serine/Threonine protein kinases, catalyti: Y61-L316	HMMER_SMART
			Tyrosine kinase, catalytic domain: Y61-L316	HMMER_SMART
			Receptor tyrosine kinase class III IPB001824: M97-R151, Q161-A200, A218-P260	BLIMPS_BLOCKS
			Protein kinases signatures and profile: Q161-S214	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: M137-N150, H175-L193, T242-N264, F285-I307	BLIMPS_PRINTS
			PROTEIN KINASE DOMAIN DM00004 P51954|6-248: I64-I307 P51955|10-261: V63-I307 P51957|8-251: I64-I307 Q08942|22-269: I67-I307	BLAST_DOMO

2

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S53, S105, S126, S300, S399, S487, S501, S556, S574, S754, S781, S794, S804, S838, T32, T256, T442, T640, T660, T711, T763, T799, T821, T834, T865, T873, Y474, Y605	MOTIFS
			Potential Glycosylation Sites: N212, N240, N636, N861	MOTIFS
			Protein kinases ATP-binding region signature: I67-K90	MOTIFS
			Serine/Threonine protein kinases active-site signature: I181-L193	MOTIFS
3	7524408CD1	487	GDA1/CD39 (nucleoside phosphatase) family: T80-K487	HMMER_PFAM
			Cytosolic domain: M1-I34 Transmembrane domain: M35-I54 Non-cytosolic domain: R55-K487	TMHMMER
			GDA1/CD39 family of nucleoside phosphatase IPB000407: I91-Y105, P173-R183, I217-E238, G268-Y281	BLIMPS_BLOCKS
			HYDROLASE TRANSMEMBRANE PROTEIN NUCLEOSIDE CD39 NUCLEOSIDE-TRIPHOSPHATASE TRIPHOSPHATE NTP-ASE PRECURSOR ATP-DIPHOSPHOHYDROLASE PD003822:N86-K487 I91-Y105	BLAST_PRODOM
			GUANOSINE DIPHOSPHATASE-LIKE PROTEIN KIAA0392 PD070805: M1-P85	BLAST_PRODOM

3

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7526163CD1	1309	ACTIVATION; NUCLEOSIDE; ANTIGEN; LYMPHOID; DM02628 |I56242|40-471:N86-S291 K264-F472 |P49961|40-471:N86-S291 K264-F472 |P32621|84-517:Y89-G235 T266-Y435 |P40009|1-462: N84-S479	BLAST_DOMO
			Potential Phosphorylation Sites: S212, S218, S292, S479, T75, T144, T266, Y175, Y477	MOTIFS
			Potential Glycosylation Sites: N404, N407	MOTIFS
			PDZ domain (Also known as DHR or GLGF): P950-L1037	HMMER_PFAM
			Protein kinase domain: F367-F640	HMMER_PFAM
			Domain present in PSD-95, Dlg, and ZO-1/2: K958-E1038	HMMER_SMART
			Extension to Ser/Thr-type protein kinases: L641-F704	HMMER_SMART
			Serine/Threonine protein kinases, catalyt: F367-F640	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H482-L497, Y557-G567	BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: M444-K457, Y480-I498, V561-D583	BLIMPS_PRINTS
			PROTEIN KINASE SERINE/THREONINE KIN4 MICROTUBULE ASSOCIATED TESTIS SPECIFIC TESTIS SPECIFIC MAST205 PD041650: R177-D366	BLAST_PRODOM

4

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE 205KD TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE MAST205 KINASE PD069998: T1034-D1128	BLAST_PRODOM
			MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE 205KD TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE MAST205 KINASE PD135564: G11-Y176	BLAST_PRODOM
			MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE 205KD TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE MAST205 KINASE PD182663: T719-W984	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004 |S42867|75-498:I370-K513 H527-F668 A1135-A1148 |A54602|455-712: T369-G627	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM08046 |P05986|1-397:S365-K513 V540-E684 |P06244|1-396:D366-K513 V540-E684	BLAST_DOMO

5

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S57, S61, S85, S108, S146, S257, S336, S358, S365, S621, S666, S672, S680, S690, S701, S709, S726, S738, S747, S767, S792, S793, S921, S930, S946, S956, S963, S985, S1031, S1041, S1060, S1063, S1074, S1080, S1101, S1180, S1273, S1215, S1257, S1262, T1068, T1181, T121, T420, T584, T670, T685, T721, T907, T1035, T1036, T235, T369, T436, T854, T1088, T1175, T1036	MOTIFS
			Potential Glycosylation Sites: N64, N1039	MOTIFS
			Serine/Threonine protein kinases active-site signature: I486-I498	MOTIFS
5	7526158CD1	1331	PDZ domain (Also known as DHR or GLGF): P972-L1059	HMMER_PFAM
			Protein kinase domain: F389-F662	HMMER_PFAM
			Domain present in PSD-95, Dlg, and ZO-1/2: K980-E1060	HMMER_SMART
			Serine/Threonine protein kinases, catalyt: F389-F662	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H504-L519, Y579-G589	BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: M466-K479, Y502-I520, V583-D605	BLIMPS_PRINTS
			PROTEIN KINASE SERINE/THREONINE KIN4 MICROTUBULE ASSOCIATED TESTIS SPECIFIC TESTIS SPECIFIC MAST205 PD041650: R199-D388	BLAST_PRODOM

PF-1688 P

6

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE 205KD TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE MAST205 KINASE PD069998: T1056-D1150	BLAST_PRODOM
			MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE 205KD TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE MAST205 KINASE PD135564: C47-Y198	BLAST_PRODOM
			MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE 205KD TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE MAST205 KINASE PD182663: T741-W1006	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004 |S42867|75-498:I392-K535 H549-F690 A1157-A1170 |A54602|455-712: T391-G649	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM08046 |P05986|1-397:S387-K535 V562-E706 |P06244|1-396:D388-K535 V562-E706	BLAST_DOMO

PF-1688 P

7

PF-1688 P

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S3, S46, S79, S83, S107, S130, S168, S279, S358, S387, S643, S688, S694, S702, S712, S723, S731, S748, S760, S769, S789, S814, S815, S943, S952, S968, S978, S985, S1007, S1053, S1063, S1082, S1085, S1096, S1102, S1123, S1202, S1284, S1295, S1237, S1279, S1284, T1090, T1203, T143, T442, T606, T692, T707, T743, T929, T1057, T1328, T257, T391, T458, T876, T1110, T1197, T1328	MOTIFS
			Serine/Threonine protein kinases active-site signature: I508-I520	MOTIFS

8

PF-1688 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
6/2509577CB1/ 3912	3840-3912, 1-1311, 2685-3085	GBI.NEK10.edit	1	3912
		56087809H1	692	1142
		56087825H1	692	1191
		56087901H1	692	1374
		56087901J1	692	1374
		56087825J1	692	1383
		56087925H1	692	1385
		56087809J1	692	1385
		56087925J1	692	1385
		5026615F6 (COLCDIT01)	991	1569
		55065621H1	1053	1112
		2509577H1 (CONUTUT01)	1118	1449
		72335521V1	1118	1619
		71869970V1	1118	1704
		72335487V1	1118	1717
		71871205V1	1118	1774
		72648030V1	1193	1462
		72645481V1	1193	1505
		55077251H1	1193	1625
		55077255H1	1193	1667
		72647270V1	1193	1687
		72646262V1	1193	1702
		72646441V1	1193	1716
		55077256H1	1193	1724
		72646443V1	1193	1735
		72334887V1	1202	1884
		g22918955	1306	1624
		9653267U2	1308	1841
		73282635V1	1308	1866
		9657511U2	1308	1989
		6097133F8 (UTRENOT09)	1308	2001
		9657511U1	1308	2006
		73282635D1	1323	2016
		6097133F6 (UTRENOT09)	1333	1833
		55056758H1	1831	2300
		55056782H1	1831	2302
		55056710H1	1831	2302
		55056742H1	1831	2302
		55056718H1	1831	2302
		55056790H1	1831	2302
		55056766H1	1831	2302
		55056750H1	1831	2302
		55056726H1	1831	2302
		55056734H1	1831	2302

1

PF-1688 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		73281314V1	1838	2524
		9635118U1	1911	2622
		9635118U3	1921	2796
		9668815U2	2158	2951
		73282073V1	2231	2430
		g13996607	2266	2685
		9668815U1	2269	2951
		73282214V1	2359	2951
		5546336H1 (TESTNOC01)	2441	2632
		5546336F8 (TESTNOC01)	2441	2957
		5546336F7 (TESTNOC01)	2441	2965
		6999077H1 (BRAXTDR17)	2466	3026
		6999077F8 (BRAXTDR17)	2467	2943
		6999077R8 (BRAXTDR17)	2475	3193
		4123469H1 (BRSTTUT25)	2676	2925
		2392071H1 (PROSNON01)	2969	3225
7/7505222CB1/ 3229	1848-2289, 1-30, 1097-1517, 228-446	GBI.g22053291.edit4	1	2813
		GBI.g22053291.edit9	147	2813
		7469170H1 (LUNGNOE02)	1800	2249
		6322794H1 (LUNGDIN02)	1976	2224
		6322715H1 (LUNGDIN02)	1976	2224
		6322794F6 (LUNGDIN02)	1976	2590
		6322794T6 (LUNGDIN02)	2139	2331
		72551805D1	2290	2373
		72549212D1	2290	2484
		2990137H1 (KIDNFET02)	2290	2512
		72614971V1	2290	2573
		71684926V1	2290	2587
		2990137F6 (KIDNFET02)	2290	2592
		2775895F6 (PANCNOT15)	2290	2596
		g2046095	2290	2601
		8467485U1	2290	2626
		8533976U1	2290	2636
		72551755D1	2290	2641
		71529926V1	2290	2669
		71530077V1	2290	2673
		72612548V1	2290	2702
		72482031D1	2290	2732
		72617183V1	2290	2755
		72548079D1	2290	2760
		73229718D1	2290	2770
		73229718V1	2290	2770
		72483936D1	2290	2776
		72613936V1	2290	2783

2

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		71530048V1	2290	2825
		71530248V1	2290	2861
		72483157D1	2290	2864
		72551004D1	2290	2871
		71530724V1	2290	2899
		72482548D1	2290	2981
		g2051645	2295	2804
		2775895T6 (PANCNOT15)	2315	2833
		72482840D1	2427	3094
		72483342D1	2427	3215
		72548002D1	2465	3204
		72551242D1	2482	3211
		71914061V1	2495	2899
		71530960V1	2509	2900
		72550024D1	2536	3229
		71528089V1	2538	2900
		72615074V1	2575	2812
		2990137T6 (KIDNFET02)	2596	2819
		72613394V1	2649	3229
8/7524408CB1/ 2100	1620-1641	95078842J1	1	753
		95078918H1	1	877
		95078842R6	2	669
		95078982R6	2	676
		95078918F6	2	718
		GBI_NT_008130_001.8_edit	2	2099
		9792742U2	529	1448
		9792742U1	548	1422
		9792741U2	631	1515
		9792741U1	640	1476
		95078918J1	1257	2100
		95078850H1	1319	2100
		95078918R6	1340	2099
		95078842H1	1345	2100
		95078850F6	1353	2099
9/7526163CB1/ 4213	1-1638, 4190-4213, 3559-3648, 2943-3018	55114261H1	1	718
		55114385H1	1	789
		g3043645_CD	11	4017
		6889766H1 (BRAITDR03)	14	590
		6951981F8 (BRAITDR02)	17	664
		6951981H1 (BRAITDR02)	17	674
		55114385J1	24	849
		7751843J1 (HEAONOE01)	79	221

3

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		GBI.g22053494.edit1	169	4213
		55114269H1	171	849
		55114253H1	241	929
		55114377J1	241	929
		7676193J2 (NOSETUE01)	311	946
		70505656V1	380	828
		6951981R8 (BRAITDR02)	421	1058
		55114261J1	440	1069
		6977576F8 (BRAHTDR04)	466	1171
		55055940H1	494	984
		55055924H1	494	1004
		55055980H1	494	1017
		72634030V1	494	1019
		72635945V1	494	1037
		72636145V1	494	1122
		55055972J1	494	1270
		55055956J1	494	1279
		72633368V1	509	1065
		70501435V1	541	1058
		70504122V1	559	1058
		72469204D1	654	1148
		72466303D1	674	1307
		72634469V1	676	1286
		70505681V1	722	874
		72637358V1	731	1267
		72633390V1	750	1460
		72464094D1	850	1386
		72635309V1	940	1516
		8270466U1	960	1765
		72633444V1	963	1639
		72465303D1	965	1553
		72633860V1	965	1590
		72632931V1	973	1633
		72464155D1	975	1571
		72465039D1	1023	1616
		72464370D1	1071	1647
		55085554H1	1079	1569
		72467177D1	1093	1716
		55085553J1	1101	1585
		72466685D1	1110	1658
		72468522D1	1110	1689
		72463771D1	1110	1725
		72464836D1	1117	1590
		55085554J1	1120	1788
		8270465U1	1127	1992
		72464654D1	1134	1635

4

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		72467808D1	1135	1581
		72466006D1	1144	1648
		72465671D1	1145	1876
		72464963D1	1157	1598
		72468460D1	1161	1839
		72468357D1	1162	1772
		72464181D1	1196	1857
		72463830D1	1204	1762
		72635760V1	1207	1915
		72464363D1	1215	1877
		72466089D1	1262	1805
		72464123D1	1277	1775
		72468937D1	1280	1981
		72633513V1	1290	1915
		g749672	1298	1711
		72469574D1	1303	1964
		72466314D1	1349	1825
		72634480V1	1354	1973
		72637375V1	1367	2079
		72467343D1	1378	2014
		72468676D1	1384	2081
		72467479D1	1418	1946
		55076784H1	1422	1915
		72467492D1	1427	1948
		72464857D1	1431	1980
		72467007D1	1433	2054
		72463976D1	1440	2126
		72466325D1	1460	2112
		72463745D1	1488	1898
		72634757V1	1533	2187
		55055948H1	1544	2393
		55055924J1	1559	2390
		55055940J1	1570	2397
		55055972H1	1583	2386
		72466202D1	1608	2050
		72466976D1	1612	2210
		6826871J1 (SINTNOR01)	1639	2204
		g11974952	1639	2221
		g14078572	1639	2291
		g14075791	1639	2342
		72464236D1	1641	2322
		72856783D1	1666	2393
		72856783V1	1676	2347
		3268480F6 (BRAINOT20)	1684	2366
		7972002H1 (CONRTUC01)	1692	2223
		72857113V1	1701	2393

PF-1688 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		72636173V1	1704	2231
		72856562D1	1706	2320
		72856562V1	1706	2393
		55114293J1	1711	2531
		g11970048	1724	2431
		7623318J1 (HEARFEE03)	1755	2310
		72836565V1	1760	2393
		55148292J1	1762	2192
		55148276J1	1763	2192
		6767896F7 (BRAUNOR01)	1771	2398
		72468248D1	1793	2259
		72467410D1	1798	2393
		55055956H1	1809	2393
		72839279V1	1813	2393
		72468512D1	1816	2393
		55148268H1	1818	2333
		55148268J1	1818	2333
		72464180D1	1820	2393
		72841950V1	1856	2393
	\	72468080D1	1858	2393
		72472316D1	1885	2393
		55076784J1	1897	2393
		55076780J1	1913	2393
		72633391V1	1915	2393
		55076785J1	1916	2393
		55076786J1	1936	2393
		55076787J1	1937	2393
		7960593J1 (TONSDIE01)	1980	2644
		72465305D1	1993	2393
		55075185J1	2005	2393
		g12287025	2021	2484
		7710071H1 (PANCNOE02)	2029	2419
		55075187J1	2086	2393
		8247429H1 (LIVRUNF02)	2148	2792
		55148268J1	2156	2180
		55148268H1	2156	2180
		73208142D1	2276	3024
		73208142V1	2289	3020
		9468312U1	2314	3220
		g7142756	2329	2786
		g7142755	2329	2829
		7582850H1 (BRAIFEC01)	2351	2940
		7582850F6 (BRAIFEC01)	2363	3123
		g7142551	2586	3050
		7623318H1 (HEARFEE03)	2602	3014
		7960593H1 (TONSDIE01)	2628	3274

6

PF-1688 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments)	Sequence Fragments	5' Position	3' Position
		73209446D1	2696	3341
		73209614D1	2731	3244
		73210136D1	2840	3377
		73209446V1	2863	3679
		g3162737	2945	3400
		7717974H1 (SINTFEE02)	3402	3605
		g11975870	3428	3935
		6993194H1 (BRAQTDR02)	3595	4034
		8179153H1 (EYERNON01)	3627	4183
		73207920D1	3632	4061
		73207920V1	3638	4091
		g7113208	3711	4185
		g7112889	3720	4189
		7219549H1 (SPLNDIC01)	3809	4191
		7219549F9 (SPLNDIC01)	3809	4212
		1894680H1 (THP1TXT04)	3909	4160
		1894680F6 (THP1TXT04)	3909	4181
		8166636H1 (MIXDTXE01)	4059	4213
10/7526158CB1/ 5991	1-1731, 2436-4780	55114385H1	1	882
		GBI.g22053494.edit1	1	5991
		55114385J1	20	942
		GBI.g22053494.edit4	36	335
		55114285J1	66	942
		7751843J1 (HEAONOE01)	99	314
		56013762J1	133	335
		55114353H1	175	942
		55114369J1	214	942
		55114277H1	241	334
		GBI.g17455923.edit1	262	4306
		55114393H1	333	941
		55114277H1	333	942
		7676193J2 (NOSETUE01)	404	1039
		70505656V1	473	921
		7526158F6 (null)	473	932
		6951981R8 (BRAITDR02)	514	1151
		55114261J1	533	1162
		6977576F8 (BRAHTDR04)	559	1264
		55055940H1	587	1077
		55055924H1	587	1097
		55055980H1	587	1110
		72634030V1	587	1112
		72635945V1	587	1130
		72636145V1	587	1215
		55055972J1	587	1363

7

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		55055956J1	587	1372
		72633368V1	602	1158
		70501435V1	634	1151
		70504122V1	652	1151
		72469204D1	747	1241
		72464155D1	1068	1664
		55085554H1	1172	1662
		72467177D1	1186	1809
		55085553J1	1194	1678
		55085554J1	1213	1881
		8270465U1	1220	2085
		72465671D1	1238	1969
		72468460D1	1254	1932
		72463830D1	1297	1855
		72464123D1	1370	1868
		72468937D1	1373	2074
		g749672	1391	1804
		72466314D1	1442	1918
		72467343D1	1471	2107
		72468676D1	1477	2174
		72467479D1	1511	2039
		55076784H1	1515	2008
		72467492D1	1520	2041
		72464857D1	1524	2073
		72463976D1	1533	2219
		72466325D1	1553	2205
		72463745D1	1581	1991
		72634757V1	1626	2280
		55055948H1	1637	2486
		55055924J1	1652	2483
		55055940J1	1663	2490
		55055972H1	1676	2479
		72466202D1	1701	2143
		72466976D1	1705	2303
		6826871J1 (SINTNOR01)	1732	2297
		g11974952	1732	2314
		g14078572	1732	2384
		g14075791	1732	2435
		72464236D1	1734	2415
		72856783D1	1759	2486
		72856783V1	1769	2440
		3268480F6 (BRAINOT20)	1777	2459
		7972002H1 (CONRTUC01)	1785	2316
		72857113V1	1794	2486
		72636173V1	1797	2324
		72856562D1	1799	2413

8

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		72856562V1	1799	2486
		55114293J1	1804	2624
		g11970048	1817	2524
		7623318J1 (HEARFEE03)	1848	2403
		72836565V1	1853	2486
		6767896F7 (BRAUNOR01)	1864	2491
		72468248D1	1886	2352
		72467410D1	1891	2486
		55055956H1	1902	2486
		72839279V1	1906	2486
		72468512D1	1909	2486
		72464180D1	1913	2486
		72841950V1	1949	2486
		72468080D1	1951	2486
		72472316D1	1978	2486
		55076784J1	1990	2486
		55076780J1	2006	2486
		72633391V1	2008	2486
		55076785J1	2009	2486
		55076786J1	2029	2486
		55076787J1	2030	2486
		7960593J1 (TONSDIE01)	2073	2737
		72465305D1	2086	2486
		55075185J1	2098	2486
		g12287025	2114	2577
		7710071H1 (PANCNOE02)	2122	2512
		55075187J1	2179	2486
		8247429H1 (LIVRUNF02)	2241	2885
		73208142D1	2369	3117
		73208142V1	2382	3113
		9468312U1	2407	3313
		g7142756	2422	2879
		g7142755	2422	2922
		7582850H1 (BRAIFEC01)	2444	3033
		7582850F6 (BRAIFEC01)	2456	3216
		g7142551	2679	3143
		7623318H1 (HEARFEE03)	2695	3107
		7960593H1 (TONSDIE01)	2721	3367
		73209446D1	2789	3434
		73209614D1	2824	3337
		73210136D1	2933	3470
		73209446V1	2956	3772
		g3162737	3038	3493
		7717974H1 (SINTFEE02)	3495	3698
		g11975870	3521	4028
		8179153H1 (EYERNON01)	3720	4276

PF-1688 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		73207920D1	3725	4154
		73207920V1	3731	4184
		6993194H1 (BRAQTDR02)	3740	4127
		g7113208	3804	4278
		g7112889	3813	4282
		7219549H1 (SPLNDIC01)	3902	4289
		7219549F9 (SPLNDIC01)	3902	4305
		1894680H1 (THP1TXT04)	4002	4253
		1894680F6 (THP1TXT04)	4002	4274
		8166636H1 (MIXDTXE01)	4152	4306
		GBI.g22053494.edit1	4403	4463
		GBI.g22053494.edit1	4537	4598

10

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library		
6	2509577CB1	TESTNOC01		
7	7505222CB1	LUNGDIN02		
9	7526163CB1	BRAITDR02		
10	7526158CB1	THP1TXT04		

1

PF-1688 P

Table 6

Library	Vector	Library Description
BRAITDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, neocortex, anterior and frontal cingulate tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
LUNGDIN02	pINCY	This normalized lung tissue library was constructed from 7.6 million independent clones from a diseased lung tissue library. Starting RNA was made from RNA isolated from diseased lung tissue. Pathology indicated ideopathic pulmonary disease. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
THP1TXT04	pINCY	Library was constructed using RNA isolated from stimulated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old male (Abbott Sample) with acute monocytic leukemia (Int. J. Cancer 26 (1980):171).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

1

4560476408.060503

PF-1688 P

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

2

PF-1688 P

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

3

PF-1688 P

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
6	2509577	5546336F7	SNP00070606	44	2484	T	T	C	V770	n/a	n/a	n/a	n/a
8	7524408	2749757H1	SNP00121108	213	1411	T	T	C	Y460	n/a	n/a	n/a	n/a
8	7524408	4739562R7	SNP00033062	268	1957	C	C	T	noncoding	n/a	n/a	n/a	n/a
10	7526158	5206370H1	SNP00130724	195	667	G	G	C	G211	n/a	n/a	n/a	n/a
10	7526158	7960593H1	SNP00071326	132	3236	C	T	C	G1067	0.46	0.37	0.67	0.46